MULTI-VIEW TIME-SERIES CLUSTERING WITH TEMPORAL STRUCTURE AND TENSOR REGULARIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We propose a tensor-based method for multi-view time-series clustering that jointly models temporal information and structure shared across views. Within each view, we combine normalized dissimilarities derived from trajectory states and first-order differences to construct a clustering objective incorporating pairwise relations and triplet geometry. View weights are learned adaptively to balance the view-specific fitting losses. Across views, we learn a common clustering embedding through symmetrized products of subspace projectors and impose tensor nuclear norm regularization on orthogonally aligned view-specific embeddings. Orthogonal alignment accommodates differences in embedding bases while preserving the subspace of each view. The final cluster assignments are obtained from the common embedding. To solve the resulting nonconvex, nonsmooth problem, we develop a two-level algorithm combining proximal majorization–minimization with an inexact augmented Lagrangian method. Under the stated conditions, we prove that the splitting constraint residual converges to zero and every accumulation point of the outer primal iterates is a first-order stationary point of the original problem. Comparative experiments on multiple datasets demonstrate the method's effectiveness in clustering multi-view time series. Parameter sensitivity analyses and ablation studies further assess the effects of key parameters and individual components on clustering performance.

1 INTRODUCTION

Multi-view time-series data describe the same objects through trajectories from different sources or feature groups. For object i in view v, the distinction from static multi-view data is

$$\begin{aligned} \text{Static:} \quad & x_i^{(v)} \in \mathbb{R}^{d_v}, \\ \text{Dynamic:} \quad & X_i^{(v)} = \left[x_i^{(v)}(1), \ldots, x_i^{(v)}(T_v)\right] \in \mathbb{R}^{d_v \times T_v}, \end{aligned}$$

where d_v and T_v denote the feature dimension and sequence length of view v. Examples include multi-sensor activity data (Yao et al., 2017), optical and radar satellite sequences (Sainte Fare Garnot et al., 2022), and high-performance computing monitoring streams (Netti et al., 2021). The goal is to group objects without labels using their complete multi-view trajectories. Discriminative information can lie in observation levels, temporal changes, and complementary information across views.

Multi-view clustering combines views through spectral co-regularization, shared low-rank structures, and subspace learning (Kumar et al., 2011; Xia et al., 2014; Gao et al., 2015). Time-series research has developed shape-based clustering and differentiable alignment losses (Paparrizos & Gravano, 2015; Cuturi & Blondel, 2017), as well as convolutional, Transformer, and contrastive representation learning (Franceschi et al., 2019; Zerveas et al., 2021; Yue et al., 2022). These approaches provide foundations for view fusion and temporal modeling. We study how temporal changes can directly shape clustering geometry while shared structure is learned across views.

Two modeling issues are central. First, temporal averaging may obscure discriminative fluctuations, while flattening retains observations without explicitly distinguishing discrepancies in states from those in first differences. Pairwise distances can also be complemented by local triplet geometry. Second, view-specific embeddings may use different coordinate bases. Applying a tensor penalty to

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

their direct stack makes regularization depend on these basis choices. Our formulation accounts for both subspace agreement and the coordinates used to represent each subspace.

We propose DTMVC, a dynamic multi-view clustering method that jointly models temporal information, within-view geometry, and shared cross-view structure. Cluster assignments are obtained from a common embedding. Experiments on six real-world datasets evaluate clustering performance, parameter sensitivity, and the effects of individual components. Our contributions are:

- We combine normalized state and first-difference dissimilarities to construct pairwise and sparse triadic clustering losses, with adaptive weights balancing the view-specific fitting energies.
- We couple projection-product consensus with tensor nuclear norm regularization of orthogonally aligned embeddings. Alignment accommodates differences in embedding bases while preserving each view's subspace projector.
- We develop a two-level proximal MM–iALM algorithm for the resulting nonconvex, nonsmooth problem. Under the stated conditions, the splitting residual vanishes, and every primal accumulation point admits a multiplier satisfying the split KKT conditions and corresponds to a first-order stationary point of the original problem.

2 Related Work

Multi-view clustering. Low-rank tensor methods couple view-specific self-representations (Zhang et al., 2015; Xie et al., 2018) or decompose tensors of transition matrices to extract shared structure (Wu et al., 2019). Joint graph and tensor learning further allows affinity construction and cross-view regularization to inform each other (Wu et al., 2020). Recent scalable approaches operate on compact representations: S^2MVTC smooths embedding tensors (Long et al., 2024), DSTL disentangles semantic-related and semantic-unrelated representations through slim tensor learning (Xu et al., 2025), and LMTC uses implicit linear kernels to model sample similarities efficiently (Liu et al., 2025). Beyond pairwise sample relations, hypergraph spectral learning models relations among multiple objects (Zhou et al., 2006), while Cai et al. (2024) combine multi-order affinities with co-regularized multiview embeddings. Projection products offer a complementary approach to identifying joint subspaces (Sergazinov et al., 2025). These studies motivate our treatment of local sample geometry and shared subspaces. DTMVC constructs this geometry from trajectory states and first differences, and combines projection-product consensus with a tensor penalty on orthogonally aligned embeddings.

Time-series representation and clustering. DTCR integrates temporal reconstruction and a k-means objective to learn cluster-specific sequence representations (Ma et al., 2019). Temporal neighborhood coding learns representations from local temporal neighborhoods (Tonekaboni et al., 2021), and TS-TCC combines temporal and contextual contrasting across augmented sequences (Eldele et al., 2021). CSL learns shapelet representations through multi-grained contrasting and multi-scale alignment (Liang et al., 2023). More recently, EMTC uses adaptive masking and cluster-guided contrastive learning to reduce the influence of redundant observations (Tan et al., 2026). MVCIMTS treats individual variables as views and jointly addresses missing-information recovery and clustering (Li et al., 2025). Our setting consists of complete, sample-matched trajectories whose views may contain multiple features. We explicitly distinguish discrepancies in observation levels from those in adjacent changes, use both to define pairwise and sparse triadic clustering losses, and learn the view embeddings together with their shared clustering representation.

3 Proposed Method

Notation. We use $A^\top$, $\mathrm{Tr}(A)$, and I_c for the transpose, trace, and $c\times c$ identity matrix, respectively. The symbols $\langle\cdot,\cdot\rangle$ and $\|\cdot\|_F$ denote the Frobenius inner product and norm. For a vector a, $a^{\otimes 3} = a\otimes a\otimes a$ denotes its third-order outer product. Other notation is defined at first use.

Figure 1 outlines DTMVC, which partitions n complete, sample-matched multi-view trajectories into $2 \le c < n$ clusters. Trajectories share a temporal grid within each view.

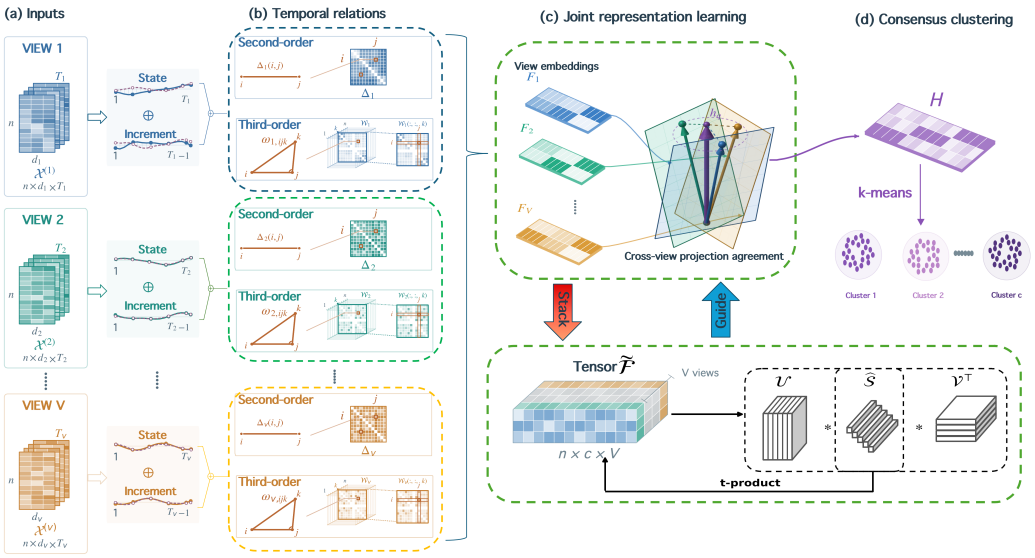

Figure 1: Overview of the proposed DTMVC framework.

3.1 Clustering with Temporal Structure

Within each view, we standardize each feature using its mean and standard deviation over all samples and time points, and remove features with zero variance. For the standardized observations $Z^{(v)}$, with $T_v \geq 2$, set $Z^{(v,0)} = Z^{(v)}$ and $Z^{(v,1)}_{ift} = Z^{(v)}_{if,t+1} - Z^{(v)}_{ift}$. The state component compares levels at corresponding time points; the first-difference component makes discrepancies in adjacent changes explicit. For $q \in \{0, 1\}$, define

$$d_{q,v,ij} = \sum_{f=1}^{d_v} \left[\frac{1}{T_v - q} \sum_{t=1}^{T_v - q} \left(Z^{(v,q)}_{ift} - Z^{(v,q)}_{jft} \right)^2 \right]^{1/2}, \qquad \bar{d}_{q,v} = \frac{2 \sum_{i<j} d_{q,v,ij}}{n(n-1)}, \tag{1}$$

where d_v counts the retained features. We normalize each component by its off-diagonal mean so that their relative contribution is not determined by their scales. After discarding views with $\bar{d}_{0,v} = 0$, let $V \geq 2$ be the number of retained views and set

$$(\Delta_v)_{ij} = \begin{cases} \dfrac{1}{2} \left(\dfrac{d_{0,v,ij}}{\bar{d}_{0,v}} + \dfrac{d_{1,v,ij}}{\bar{d}_{1,v}} \right), & \bar{d}_{1,v} > 0, \\ \dfrac{d_{0,v,ij}}{\bar{d}_{0,v}}, & \bar{d}_{1,v} = 0. \end{cases} \tag{2}$$

Each Δ_v is symmetric and has unit off-diagonal mean.

For normalized cluster indicators, the criterion $\frac{1}{2} \operatorname{Tr}(F_v^\top \Delta_v F_v)$ sums within-cluster dissimilarities weighted by inverse cluster sizes. We retain the full matrix Δ_v and relax the indicators to an embedding $F_v = [f_{v,1}, \ldots, f_{v,c}] \in \mathbb{R}^{n \times c}$ satisfying $F_v^\top F_v = I_c$. Subtracting the minimum of this relaxed criterion gives

$$\mathcal{L}_{2,v}(F_v) = \tfrac{1}{2} \operatorname{Tr}(F_v^\top \Delta_v F_v) - b_{2,v}, \qquad b_{2,v} = \tfrac{1}{2} \sum_{j=1}^{c} \lambda_j^{\uparrow}(\Delta_v). \tag{3}$$

The Ky–Fan variational principle ensures $\mathcal{L}_{2,v} \geq 0$.

The pairwise loss uses dissimilarity magnitudes. To also capture triangle shapes, we introduce an angular triadic score inspired by Cai et al. (2024). For each sample i in view v, let $\mathcal{N}_v(i)$ contain its $k_{\mathrm{nn}} = \min\{10, n-1\}$ nearest other samples under Δ_v, with ties resolved once during preprocessing. We use the view-specific support $\mathcal{T}_v = \bigcup_{i=1}^{n} \{\{i, j, k\} : j, k \in \mathcal{N}_v(i),\ j < k\}$. Triples generated by multiple anchors are included only once.

For $e = \{i, j, k\} \in \mathcal{T}_v$ with positive side lengths $(\Delta_v)_{ij}$, $(\Delta_v)_{ik}$, and $(\Delta_v)_{jk}$, let ψ_i, ψ_j, ψ_k be its comparison-triangle angles. Set $\omega_{v,e} = (3 - \cos\psi_i - \cos\psi_j - \cos\psi_k)/6$. This weight depends on triangle shape and is unchanged by a common rescaling of the three sides. Triples with a zero side receive weight zero; positive collinear triples are retained.

Assign $\omega_{v,e}$ to all six permutations of each $e \in \mathcal{T}_v$ in $\mathcal{A}_v \in \mathbb{R}^{n\times n\times n}$, and set all other entries to zero. Thus $\|\mathcal{A}_v\|_F^2 = 6\sum_{e\in\mathcal{T}_v} \omega_{v,e}^2$. Set $\mathcal{W}_v = \mathcal{A}_v/\|\mathcal{A}_v\|_F$ when $\mathcal{A}_v \neq 0$, and $\mathcal{W}_v = 0$ otherwise. Define

$$\begin{aligned} \Phi_{3,v}(F_v) &= \left\langle \mathcal{W}_v, \sum_{\ell=1}^{c} f_{v,\ell}^{\otimes 3} \right\rangle, \\ \mathcal{L}_{3,v}(F_v) &= \sqrt{c}\,\|\mathcal{W}_v\|_F - \Phi_{3,v}(F_v). \end{aligned} \tag{4}$$

Since $\|\sum_{\ell=1}^{c} f_{v,\ell}^{\otimes 3}\|_F = \sqrt{c}$, Cauchy–Schwarz gives $\mathcal{L}_{3,v} \geq 0$. We store only unordered triples and evaluate their contributions without forming a dense tensor. Both Δ_v and $\mathcal{W}_v$ are computed once from the observations and remain fixed during optimization.

Let $E_v(F_v) = \mathcal{L}_{2,v}(F_v) + \beta\mathcal{L}_{3,v}(F_v)$, where $\beta \geq 0$ controls the triadic loss. We jointly learn the embeddings and weights through $\sum_v \alpha_v^r E_v(F_v)$, with $\alpha_v \geq 0$, $\sum_v \alpha_v = 1$, and $r > 1$ controlling weight concentration. The constants in both losses are retained because they affect the optimization of $\boldsymbol{\alpha}$.

3.2 Consensus through Projection Products

The weighted fitting losses do not directly couple the view subspaces. Let $H \in \mathbb{R}^{n\times c}$ satisfy $H^\top H = I_c$ and $P_v = F_v F_v^\top$. Motivated by projection products in shared subspace learning (Sergazinov et al., 2025), define

$$M_\times = \frac{1}{2N_p}\sum_{u<v}(P_u P_v + P_v P_u), \qquad N_p = \frac{V(V-1)}{2}. \tag{5}$$

Since $\mathrm{Tr}(H^\top M_\times H) = N_p^{-1}\sum_{u<v}\langle P_u H, P_v H\rangle_F$, we add $-\gamma\,\mathrm{Tr}(H^\top M_\times H)$, with $\gamma \geq 0$, to encourage agreement between these projections. All view pairs contribute equally; the adaptive weights apply only to the fitting energies.

3.3 Aligned Tensor Regularization

Direct stacking makes the tensor penalty depend on view-specific embedding bases. We therefore form $\widetilde{\mathcal{F}} \in \mathbb{R}^{n\times c\times V}$ with $\widetilde{\mathcal{F}}(:,:,v) = F_v R_v$, where $R_v \in \mathbb{R}^{c\times c}$ and $R_v^\top R_v = I_c$. The alignment acts only in the tensor term and preserves $(F_v R_v)(F_v R_v)^\top = P_v$. The tensor nuclear norm penalizes singular values of the Fourier slices along the view mode. Combining these terms gives

$$\begin{aligned} \min_{\{F_v,R_v\}_{v=1}^V,\, H,\, \boldsymbol{\alpha}} \quad & \sum_{v=1}^{V} \alpha_v^r E_v(F_v) - \gamma\,\mathrm{Tr}(H^\top M_\times H) + \lambda\|\widetilde{\mathcal{F}}\|_{\mathrm{TNN}} \\ \text{s.t.} \quad & F_v^\top F_v = R_v^\top R_v = I_c, \quad H^\top H = I_c, \\ & \alpha_v \geq 0, \quad \sum_{v=1}^{V}\alpha_v = 1, \quad \widetilde{\mathcal{F}}(:,:,v) = F_v R_v \quad (\forall v), \end{aligned} \tag{6}$$

Here $\lambda \geq 0$ controls tensor regularization. Cluster labels are obtained by applying k-means with $k = c$ to the row-normalized H, leaving zero rows unchanged.

4 Optimization

To separate the tensor nuclear norm from the orthogonality constraints, we introduce $\mathcal{J} = \widetilde{\mathcal{F}}$ and retain the remaining constraints in (6). Let $\mathcal{Q}$ denote its objective with $\|\widetilde{\mathcal{F}}\|_{\mathrm{TNN}}$ replaced by $\|\mathcal{J}\|_{\mathrm{TNN}}$. At outer iteration s, the augmented Lagrangian is

$$L_{\rho_s} = \mathcal{Q} + \langle \mathcal{Y}^s, \widetilde{\mathcal{F}} - \mathcal{J}\rangle + \frac{\rho_s}{2}\|\widetilde{\mathcal{F}} - \mathcal{J}\|_F^2. \tag{7}$$

We use a two-level MM–iALM scheme to seek a KKT point of the split problem. With $\mathcal{U}^s = \mathcal{Y}^s/\rho_s$, the inner loop controls stationarity error at fixed $(\mathcal{Y}^s, \rho_s)$, while the outer loop updates the multiplier and penalty to enforce $\widetilde{\mathcal{F}} = \mathcal{J}$.

4.1 Block Updates

Each inner sweep follows the order $\{F_v\}_{v=1}^V \to \{R_v\}_{v=1}^V \to H \to \boldsymbol{\alpha} \to \mathcal{J}$. Below, t denotes the inner index and s is suppressed on the primal variables. For a thin SVD $A = U\Sigma V^\top$, write $\mathrm{Polar}(A) = UV^\top$.

Updating F_v. Set $S_{-v}^t = \sum_{u<v} P_u^{t+1} + \sum_{u>v} P_u^t$, $P_H^t = H^t(H^t)^\top$, and $B_v^t = \mathcal{J}_v^t - \mathcal{U}_v^s$, where tensor subscripts denote frontal slices. Since $\|FR_v^t\|_F^2 = c$ on the constraint set, the reduced block gradient is

$$G_v^t = (\alpha_v^t)^r \nabla E_v(F_v^t) - \frac{\gamma}{N_p}(P_H^t S_{-v}^t + S_{-v}^t P_H^t)F_v^t - \rho_s B_v^t (R_v^t)^\top, \tag{8}$$

with $[\nabla E_v(F)]_{iq} = (\Delta_v F)_{iq} - 3\beta \sum_{j,k} (\mathcal{W}_v)_{ijk} F_{jq} F_{kq}$. The cubic term is handled by a quadratic MM surrogate. A valid curvature bound is $\ell_{F,v}^t = (\alpha_v^t)^r(\|\Delta_v\|_\infty + 6\beta\|\mathcal{W}_v\|_F) + 4\gamma/V$, where $\|\Delta_v\|_\infty$ is the maximum absolute row-sum norm. For $\epsilon_\tau > 0$, the surrogate is minimized by

$$\tau_v^t = \ell_{F,v}^t + \epsilon_\tau \max\{1, \ell_{F,v}^t\}, \qquad F_v^{t+1} = \mathrm{Polar}(\tau_v^t F_v^t - G_v^t). \tag{9}$$

Updating R_v and H. Using the proximal penalties $\frac{\eta_R}{2}\|R - R_v^t\|_F^2$ and $\frac{\eta_H}{2}\|HH^\top - P_H^t\|_F^2$, with $\eta_R, \eta_H > 0$, gives

$$R_v^{t+1} = \mathrm{Polar}\big(\rho_s (F_v^{t+1})^\top B_v^t + \eta_R R_v^t\big), \tag{10}$$

$$H^{t+1} = \mathrm{TopEig}_c\big(\gamma M_\times^{t+1} + \eta_H P_H^t\big). \tag{11}$$

Here $M_\times^{t+1}$ uses the updated view embeddings, and TopEig_c returns orthonormal eigenvectors for the c largest algebraic eigenvalues.

Updating $\boldsymbol{\alpha}$. Set $E_v^{t+1} = E_v(F_v^{t+1})$. If all energies are positive, the minimizer is

$$\alpha_v^{t+1} = \frac{(E_v^{t+1})^{-1/(r-1)}}{\sum_{u=1}^V (E_u^{t+1})^{-1/(r-1)}}. \tag{12}$$

Otherwise, let $\mathcal{Z}^{t+1} = \{v : E_v^{t+1} = 0\}$ and set $\alpha_v^{t+1} = 1/|\mathcal{Z}^{t+1}|$ for $v \in \mathcal{Z}^{t+1}$, and zero elsewhere.

Updating $\mathcal{J}$. With $\widetilde{\mathcal{F}}^{t+1}(:,:,v) = F_v^{t+1} R_v^{t+1}$, update

$$\mathcal{J}^{t+1} = \mathrm{tSVT}_{\lambda/\rho_s}\big(\widetilde{\mathcal{F}}^{t+1} + \mathcal{U}^s\big). \tag{13}$$

Here tSVT thresholds the singular values of the third-mode Fourier slices at λ/ρ_s and applies the inverse transform. Block-update derivations are given in Appendix A.

4.2 Two-Level MM–iALM Scheme

After a complete inner sweep, form the candidate multiplier

$$\widehat{\mathcal{Y}}^{s,t+1} = \mathcal{Y}^s + \rho_s\big(\widetilde{\mathcal{F}}^{s,t+1} - \mathcal{J}^{s,t+1}\big). \tag{14}$$

Let $L = \mathcal{Q} + \langle \mathcal{Y}, \widetilde{\mathcal{F}} - \mathcal{J} \rangle$ be the ordinary Lagrangian. At the completed primal point, hold $\mathcal{Y} = \widehat{\mathcal{Y}}^{s,t+1}$ fixed and evaluate

$$\mathcal{S}_s^{t+1} = \max\left\{ \max_v \|\operatorname{grad}_{F_v} L\|_F, \max_v \|\operatorname{grad}_{R_v} L\|_F, \|\operatorname{grad}_H L\|_F \right\}. \tag{15}$$

Here $\operatorname{grad}_X L = \nabla_X L - X\,\mathrm{sym}(X^\top \nabla_X L)$, with $\mathrm{sym}(A) = (A + A^\top)/2$. The exact weight and final tensor updates satisfy their block optimality conditions; in particular, $\widehat{\mathcal{Y}}^{s,t+1} \in \lambda \partial \|\mathcal{J}^{s,t+1}\|_{\mathrm{TNN}}$. The proximal penalties used in the updates are excluded from this test.

Algorithm 1 DTMVC via Two-Level Proximal MM–iALM

Input: Multi-view trajectories, cluster count c, and model and solver parameters.
Output: Common embedding H and labels $\widehat{\boldsymbol{y}}$.

Preprocess the views and construct $\{\Delta_v, \mathcal{W}_v, b_{2,v}\}_{v=1}^{V}$ as in Section 3.1.
Initialize $F_v^0, R_v^0, H^0, \boldsymbol{\alpha}^0$ feasibly; set $\mathcal{J}^0 = \widetilde{\mathcal{F}}^0$, $\mathcal{Y}^0 = 0$, and $s = 0$.
repeat
  Fix $(\mathcal{Y}^s, \rho_s)$; set $\mathcal{U}^s = \mathcal{Y}^s/\rho_s$, $\varepsilon_s = \varepsilon_0 \xi^s$, and $t = 0$. Initialize the inner variables from outer iterate s.
  repeat
    Update F_v sequentially, then $\{R_v\}, H, \boldsymbol{\alpha}, \mathcal{J}$ by (9)–(13), including the zero-energy weight rule.
    Compute $\widehat{\mathcal{Y}}^{s,t+1}$ and $\mathcal{S}_s^{t+1}$ by (14)–(15); set $t \leftarrow t + 1$.
  until $\mathcal{S}_s^t \le \varepsilon_s$
  Set $t_s = t$; accept the inner point as outer iterate $s + 1$.
  Set $\mathcal{Y}^{s+1} = \widehat{\mathcal{Y}}^{s,t_s}$ and compute r_{s+1}.
  Set $\rho_{s+1} = \rho_0$ if $s = 0$; otherwise use (16). Recompute $\mathcal{U}^{s+1} = \mathcal{Y}^{s+1}/\rho_{s+1}$.
  Set $s \leftarrow s + 1$.
until $r_s \le \varepsilon_{\text{feas}}$ and $\mathcal{S}_{s-1}^{t_{s-1}} \le \varepsilon_{\text{stat}}$
Set $H = H^s$ and $\bar{H} = H$; normalize nonzero rows of $\bar{H}$ to unit Euclidean norm, leaving zero rows unchanged.
Obtain $\widehat{\boldsymbol{y}}$ by k-means on the rows of $\bar{H}$ with $k = c$; **return** $H, \widehat{\boldsymbol{y}}$.

Accept the inner point when $\mathcal{S}_s^{t+1} \le \varepsilon_s$, where $\varepsilon_s = \varepsilon_0 \xi^s$, $\varepsilon_0 > 0$, and $0 < \xi < 1$. Otherwise, continue with the same $(\mathcal{Y}^s, \rho_s)$; the candidate multiplier is not fed back into the inner updates. For the accepted index t_s, promote the primal variables to outer iterate $s + 1$, set $\mathcal{Y}^{s+1} = \widehat{\mathcal{Y}}^{s,t_s}$, and compute $r_{s+1} = \|\widetilde{\mathcal{F}}^{s+1} - \mathcal{J}^{s+1}\|_F$. Set $\rho_1 = \rho_0$. For $s \ge 1$, use

$$\rho_{s+1} = \begin{cases} \rho_s, & r_{s+1} \le \theta r_s, \\ \kappa \rho_s, & r_{s+1} > \theta r_s, \end{cases} \qquad 0 < \theta < 1, \quad \kappa > 1, \tag{16}$$

and recompute $\mathcal{U}^{s+1} = \mathcal{Y}^{s+1}/\rho_{s+1}$. For prescribed positive tolerances, the outer loop stops when both $r_{s+1} \le \varepsilon_{\text{feas}}$ and $\mathcal{S}_s^{t_s} \le \varepsilon_{\text{stat}}$. Algorithm 1 summarizes the procedure.

4.3 Convergence and Computational Cost

Theorem 1. *Let Δ_v and the symmetric sparse tensors $\mathcal{W}_v$ constructed in Section 3.1 be fixed throughout optimization. Let $r > 1$, $\beta, \gamma, \lambda \ge 0$, and $\rho_0, \eta_R, \eta_H, \epsilon_\tau > 0$. Use the initialization, acceptance criterion, and penalty rule specified in Algorithm 1. Suppose all prescribed MM and proximal block updates are computed exactly. Then every inner loop terminates after finitely many sweeps. If infinitely many outer iterates are generated, the accepted primal–dual sequence is bounded and $\|\widetilde{\mathcal{F}}^{s+1} - \mathcal{J}^{s+1}\|_F \to 0$. Each primal accumulation point admits a multiplier satisfying the split KKT conditions and, after eliminating $\mathcal{J}$, is a first-order stationary point of* (6).

Appendix B gives the proof. Let $M_v = |\mathcal{T}_v|$. Since $M_v \le n\binom{k_{\text{nn}}}{2}$, the triplet lists require $O(\sum_v M_v)$ storage, and evaluating the cubic scores and their gradients costs $O(c \sum_v M_v)$ per sweep. The full pairwise matrices additionally require $O(Vn^2)$ storage and $O(Vn^2 c)$ work per sweep for their products with the embeddings.

5 Experiments

5.1 Experimental Settings

Datasets: We evaluate DTMVC on six datasets from diverse application domains: MHEALTH (Banos et al., 2014), DSADS (Barshan & Altun, 2010), Tactile12 (Lima et al., 2023), PASTIS-R (Sainte Fare Garnot & Landrieu, 2021; Sainte Fare Garnot et al., 2022), HPC-ODA (Netti

Table 1: Characteristics of the six experimental datasets.

Dataset	n	V	K	(d_v)	(T_v)
MHEALTH	1,455	3	12	$(3, 9, 9)$	125
DSADS	9,120	5	19	$(9, 9, 9, 9, 9)$	125
Tactile12	1,200	3	12	$(6, 6, 3)$	60
PASTIS-R	2,000	3	3	$(4, 4, 20)$	12
HPC-ODA	2,586	5	6	$(6, 6, 4, 7, 4)$	100
NinaPro DB5-S1	312	3	52	$(8, 8, 22)$	$(400, 320, 240)$

et al., 2021), and NinaPro DB5-S1 (Pizzolato et al., 2017). Dataset statistics are summarized in Table 1; sources and view descriptions are provided in the supplementary material.

Baselines: We compare DTMVC with six baselines to assess its clustering performance against static multi-view fusion and time-series representation learning. For the static methods, $\mathrm{S}^2\mathrm{MVTC}$ (Long et al., 2024), DSTL (Xu et al., 2025), and LMTC (Liu et al., 2025), we flatten each sample's complete sequence within each view, reshaping $X^{(v)} \in \mathbb{R}^{n\times d_v\times T_v}$ into $Z^{(v)} \in \mathbb{R}^{n\times(d_vT_v)}$ while retaining the view partition. For the time-series baselines, CSL (Liang et al., 2023) concatenates equal-length views along the channel dimension, or encodes unequal-length views separately and concatenates their representations; EMTC (Tan et al., 2026) receives time-aligned, channel-concatenated sequences; and MVCIMTS (Li et al., 2025) is adapted to encode each data view separately and concatenate the resulting representations. The latter three methods apply k-means to their learned representations. For all methods, the number of clusters equals the number of ground-truth classes.

Evaluation Metrics: We report accuracy (ACC), normalized mutual information (NMI), purity (PUR), adjusted Rand index (ARI), and F-score (F), all as percentages with higher values indicating better performance. For DTMVC, we report the mean and standard deviation over ten independently initialized k-means runs.

Parameter Settings: We evaluate $\gamma, \lambda \in \{0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30\}$ and $\beta \in \{0.001, 0.01, 0.1, 1, 4, 16, 64, 128\}$, giving 512 configurations. We fix $r = 2$ and $k_{\mathrm{nn}} = \min\{10, n-1\}$, with at most 50 inner iterations per outer iteration.

5.2 Clustering Performance

For the comparisons in Table 2, we fix $\beta = 1$, $\gamma = 10$, and $\lambda = 0.01$ across all six datasets. DTMVC achieves the highest mean ACC, NMI, PUR, ARI, and F-score on every dataset, ranking first in all 30 dataset–metric comparisons. Its average ACC gain over the strongest baseline on each dataset is 10.36 percentage points. The largest ACC gains over the strongest baselines are 25.28 and 20.87 percentage points on HPC-ODA and DSADS, respectively.

5.3 Parameter Sensitivity

Figure 2 shows, for each (γ, λ) pair, the highest mean ACC over β (bar height) and the maximizing β (color). The lower grid shows all eight β results on a shared ACC scale within each dataset; the inset enlarges the highlighted profile.

Parameter sensitivity varies across datasets. After maximizing over β, ACC varies by only 1.85 and 2.24 percentage points on PASTIS-R and NinaPro DB5-S1, respectively, compared with 19.00 points on HPC-ODA. Preferred β values also differ across datasets. We recommend starting from the main-experiment configuration, tuning β first, and jointly refining γ and λ on more sensitive datasets.

5.4 Ablation Study

We consider seven variants: (i) *w/o Differences*, using only state distances, $(\Delta_v)_{ij} = d_{0,v,ij}/\bar{d}_{0,v}$; (ii) *w/o Triadic*, removing the triadic loss by setting $\beta = 0$; (iii) *w/o Projection Product*, replacing $M_\times$ with $V^{-1}\sum_v F_v F_v^\top$; (iv) *w/o Tensor*, setting $\lambda = 0$; (v) *w/o Alignment*, fixing $R_v = I_c$; (vi) *Single-sweep*, performing one inner sweep per outer iteration; and (vii) *Mean-pooling*, replac-

Table 2: Clustering performance (%, mean ± SD). Bold and underline mark the best and second best results.

Dataset	S^2MVTC	DSTL	LMTC	CSL+ K-means	EMTC	MVCIMTS+ K-means	Our Method
ACC							
MHEALTH	14.36±0.45	38.79±3.81	42.48±3.33	49.58±3.36	26.70±1.10	51.00±3.16	**53.71±3.13**
DSADS	11.32±0.33	36.99±2.15	40.62±2.66	47.25±2.78	27.17±0.72	49.50±3.16	**70.37±4.47**
Tactile12	88.69±4.79	77.04±9.22	80.23±8.48	82.84±8.64	42.79±2.36	84.00±7.07	**91.13±4.90**
PASTIS-R	34.75±0.20	56.23±0.02	91.25±0.01	84.63±12.80	48.89±1.45	87.50±7.75	**93.51±0.04**
HPC-ODA	19.44±0.27	56.58±3.07	54.43±5.31	50.13±5.30	32.64±1.11	66.82±9.15	**92.10±10.49**
NinaPro DB5-S1	64.80±2.92	61.93±2.76	58.17±3.32	48.47±2.19	32.08±1.14	65.00±3.16	**73.60±3.01**
NMI							
MHEALTH	3.66±0.26	51.94±1.53	52.33±2.23	58.96±1.37	25.77±0.75	59.00±1.73	**64.57±2.41**
DSADS	5.83±0.29	55.52±1.18	56.91±2.17	57.91±1.26	37.99±0.35	61.50±2.00	**81.02±1.81**
Tactile12	91.73±1.81	89.52±4.13	92.80±2.81	94.11±3.21	65.79±1.15	92.00±2.65	**96.52±1.41**
PASTIS-R	0.08±0.04	12.81±0.03	74.55±0.04	62.34±17.26	10.50±0.96	68.00±10.00	**76.76±0.14**
HPC-ODA	0.55±0.06	51.09±1.77	50.14±4.62	44.07±3.73	17.96±0.46	69.06±5.83	**95.39±5.83**
NinaPro DB5-S1	87.04±0.87	84.56±0.98	79.40±1.81	73.22±1.03	62.83±0.54	85.00±1.00	**90.45±0.76**
PUR							
MHEALTH	15.36±0.49	43.63±3.19	45.80±2.45	53.13±2.43	30.71±1.10	55.00±3.00	**56.08±2.94**
DSADS	12.23±0.31	42.08±1.11	44.69±2.39	49.49±2.64	28.72±0.65	53.00±3.00	**73.94±3.54**
Tactile12	90.27±3.80	79.98±7.85	85.71±5.82	87.68±6.29	52.86±1.01	87.00±5.48	**93.10±3.53**
PASTIS-R	34.78±0.22	56.23±0.02	91.25±0.01	84.73±12.57	50.64±0.50	88.00±7.75	**93.51±0.04**
HPC-ODA	23.71±0.37	60.33±2.64	57.29±4.58	55.05±3.79	35.99±1.00	70.65±6.91	**94.56±7.00**
NinaPro DB5-S1	69.95±2.23	67.29±2.29	61.76±3.04	51.67±2.11	34.56±1.00	69.00±2.83	**77.27±2.38**
ARI							
MHEALTH	1.00±0.14	32.46±2.19	31.35±2.76	38.86±2.08	11.64±0.82	40.00±2.45	**42.42±3.00**
DSADS	0.62±0.10	21.13±2.82	30.66±3.39	36.88±1.98	13.64±0.51	39.00±2.65	**63.88±4.03**
Tactile12	85.06±4.58	73.20±12.27	80.71±7.75	83.79±8.62	23.90±1.56	81.00±7.75	**90.43±4.46**
PASTIS-R	-0.02±0.04	13.28±0.03	80.50±0.03	65.58±18.75	11.40±0.82	71.00±10.00	**81.39±0.11**
HPC-ODA	0.19±0.05	36.04±2.57	34.80±5.16	26.45±4.65	5.86±0.28	53.24±8.90	**90.71±11.58**
NinaPro DB5-S1	55.14±2.74	48.86±2.98	42.87±3.99	29.30±2.16	12.08±0.80	54.00±3.16	**66.46±2.80**
F-score							
MHEALTH	9.53±0.18	39.10±1.72	37.50±2.44	44.47±1.90	19.71±0.76	45.00±2.24	**47.45±2.70**
DSADS	6.02±0.10	27.11±2.30	34.92±3.02	40.38±1.86	18.89±0.28	43.00±2.45	**65.90±3.76**
Tactile12	86.35±4.16	75.93±10.77	82.49±6.95	85.29±7.77	33.83±1.32	83.00±7.07	**91.27±4.05**
PASTIS-R	33.46±0.04	42.57±0.03	87.00±0.02	77.35±12.03	43.52±2.40	80.00±8.37	**87.61±0.07**
HPC-ODA	17.18±0.06	48.55±1.78	47.33±3.83	40.31±3.32	26.79±0.20	62.09±6.58	**92.38±9.49**
NinaPro DB5-S1	55.97±2.68	49.87±2.90	43.88±3.92	30.57±2.11	13.73±0.80	55.00±3.00	**67.07±2.74**

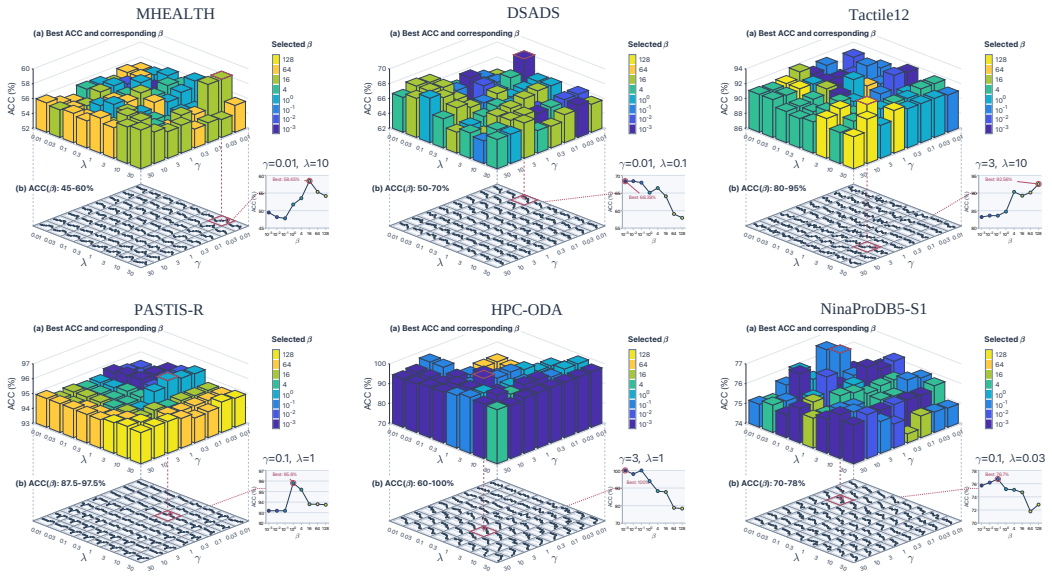

Figure 2: Sensitivity of clustering accuracy to γ, λ, and β on six datasets.

ing each trajectory with $\bar{\mathbf{x}}_i^{(v)} = T_v^{-1} \sum_t \mathbf{x}_i^{(v)}(t)$. The last variant reconstructs the geometry from standardized mean features while retaining all model terms, yielding a static multi-view clustering problem.

Table 3: Results of ablation variants and Full DTMVC on six datasets (%, mean$_{\pm \mathrm{SD}}$). Full DTMVC reports the configuration with the highest mean ACC in the grid search.

Dataset	w/o Differences	w/o Triadic	w/o Proj. Product	w/o Tensor	w/o Alignment	Single-sweep	Mean-pooling	Full DTMVC
				ACC				
MHEALTH	$44.16_{\pm 3.65}$	$51.15_{\pm 2.70}$	$48.20_{\pm 3.57}$	$46.85_{\pm 3.22}$	$50.71_{\pm 3.13}$	$47.36_{\pm 3.80}$	$48.32_{\pm 2.97}$	$58.45_{\pm 3.26}$
DSADS	$63.54_{\pm 4.46}$	$62.52_{\pm 3.46}$	$62.59_{\pm 4.08}$	$60.51_{\pm 3.43}$	$63.22_{\pm 4.24}$	$62.13_{\pm 4.51}$	$58.96_{\pm 4.56}$	$68.38_{\pm 4.99}$
Tactile12	$92.06_{\pm 4.94}$	$85.85_{\pm 5.38}$	$90.93_{\pm 4.90}$	$85.63_{\pm 5.33}$	$86.02_{\pm 5.83}$	$84.60_{\pm 4.69}$	$89.79_{\pm 5.40}$	$92.56_{\pm 4.96}$
PASTIS-R	$90.02_{\pm 10.10}$	$93.47_{\pm 0.02}$	$91.41_{\pm 8.21}$	$93.51_{\pm 0.04}$	$93.51_{\pm 0.03}$	$87.22_{\pm 0.03}$	$54.52_{\pm 7.96}$	$95.80_{\pm 0.00}$
HPC-ODA	$89.79_{\pm 9.82}$	$88.08_{\pm 10.04}$	$90.25_{\pm 10.34}$	$91.27_{\pm 9.94}$	$86.10_{\pm 10.49}$	$71.45_{\pm 9.38}$	$83.01_{\pm 6.28}$	$100.00_{\pm 0.00}$
NinaPro	$71.79_{\pm 3.04}$	$72.84_{\pm 2.97}$	$73.58_{\pm 3.18}$	$73.49_{\pm 3.65}$	$73.54_{\pm 2.89}$	$73.22_{\pm 3.13}$	$76.62_{\pm 3.25}$	$76.70_{\pm 2.87}$
				NMI				
MHEALTH	$51.00_{\pm 2.00}$	$61.84_{\pm 2.01}$	$54.14_{\pm 2.73}$	$55.05_{\pm 2.04}$	$54.57_{\pm 2.41}$	$48.93_{\pm 2.50}$	$51.21_{\pm 2.41}$	$69.18_{\pm 2.31}$
DSADS	$74.65_{\pm 2.45}$	$76.82_{\pm 1.48}$	$76.39_{\pm 1.93}$	$75.43_{\pm 1.38}$	$76.70_{\pm 1.76}$	$76.29_{\pm 1.82}$	$71.62_{\pm 3.21}$	$79.05_{\pm 1.33}$
Tactile12	$96.75_{\pm 1.35}$	$93.77_{\pm 2.28}$	$96.60_{\pm 1.49}$	$93.96_{\pm 2.42}$	$94.38_{\pm 2.30}$	$91.20_{\pm 1.57}$	$94.27_{\pm 2.09}$	$96.71_{\pm 1.21}$
PASTIS-R	$73.41_{\pm 10.52}$	$76.67_{\pm 0.11}$	$75.23_{\pm 8.76}$	$76.76_{\pm 0.14}$	$76.75_{\pm 0.12}$	$82.10_{\pm 0.06}$	$28.07_{\pm 4.79}$	$83.26_{\pm 0.01}$
HPC-ODA	$91.54_{\pm 5.59}$	$91.03_{\pm 5.46}$	$93.94_{\pm 5.73}$	$96.41_{\pm 5.39}$	$90.39_{\pm 5.83}$	$70.54_{\pm 8.84}$	$86.45_{\pm 3.53}$	$100.00_{\pm 0.00}$
NinaPro	$86.46_{\pm 0.96}$	$90.14_{\pm 0.90}$	$90.39_{\pm 0.84}$	$90.37_{\pm 1.07}$	$90.41_{\pm 0.74}$	$90.34_{\pm 0.89}$	$90.72_{\pm 1.14}$	$90.95_{\pm 0.98}$
				PUR				
MHEALTH	$47.37_{\pm 2.92}$	$54.50_{\pm 2.41}$	$50.27_{\pm 3.20}$	$50.29_{\pm 2.83}$	$52.08_{\pm 2.94}$	$49.77_{\pm 3.35}$	$50.66_{\pm 2.52}$	$60.55_{\pm 2.78}$
DSADS	$67.09_{\pm 3.71}$	$66.64_{\pm 2.90}$	$66.71_{\pm 3.30}$	$64.93_{\pm 2.90}$	$67.14_{\pm 3.29}$	$66.33_{\pm 3.64}$	$62.57_{\pm 4.06}$	$71.15_{\pm 3.40}$
Tactile12	$93.62_{\pm 3.65}$	$88.84_{\pm 3.76}$	$92.99_{\pm 3.54}$	$88.72_{\pm 3.83}$	$89.27_{\pm 4.05}$	$87.04_{\pm 3.32}$	$91.65_{\pm 3.92}$	$93.78_{\pm 3.90}$
PASTIS-R	$90.64_{\pm 7.63}$	$93.47_{\pm 0.02}$	$91.77_{\pm 6.45}$	$93.51_{\pm 0.04}$	$93.51_{\pm 0.03}$	$90.22_{\pm 0.03}$	$57.78_{\pm 6.14}$	$95.80_{\pm 0.00}$
HPC-ODA	$90.09_{\pm 6.85}$	$89.17_{\pm 6.75}$	$92.93_{\pm 7.03}$	$95.66_{\pm 6.42}$	$88.56_{\pm 7.00}$	$74.81_{\pm 8.39}$	$84.37_{\pm 4.00}$	$100.00_{\pm 0.00}$
NinaPro	$76.10_{\pm 2.49}$	$76.56_{\pm 2.41}$	$77.21_{\pm 2.57}$	$77.01_{\pm 2.97}$	$77.15_{\pm 2.34}$	$76.91_{\pm 2.51}$	$79.73_{\pm 2.76}$	$79.52_{\pm 2.20}$
				ARI				
MHEALTH	$29.93_{\pm 2.37}$	$39.91_{\pm 2.54}$	$42.53_{\pm 3.61}$	$33.72_{\pm 2.33}$	$42.42_{\pm 3.00}$	$42.77_{\pm 3.47}$	$31.99_{\pm 2.38}$	$47.82_{\pm 3.34}$
DSADS	$52.66_{\pm 4.34}$	$54.22_{\pm 2.67}$	$53.70_{\pm 3.39}$	$52.25_{\pm 2.35}$	$54.24_{\pm 3.49}$	$53.19_{\pm 3.63}$	$47.25_{\pm 5.32}$	$57.97_{\pm 3.20}$
Tactile12	$91.42_{\pm 4.05}$	$84.41_{\pm 5.76}$	$90.48_{\pm 4.83}$	$84.62_{\pm 5.88}$	$85.18_{\pm 5.70}$	$84.28_{\pm 4.77}$	$87.81_{\pm 5.10}$	$91.30_{\pm 4.11}$
PASTIS-R	$75.61_{\pm 11.90}$	$81.29_{\pm 0.06}$	$78.04_{\pm 10.09}$	$81.39_{\pm 0.11}$	$81.40_{\pm 0.08}$	$81.13_{\pm 0.07}$	$24.49_{\pm 7.79}$	$87.74_{\pm 0.00}$
HPC-ODA	$85.82_{\pm 10.43}$	$87.03_{\pm 10.91}$	$87.79_{\pm 11.51}$	$92.61_{\pm 10.98}$	$80.71_{\pm 11.58}$	$61.50_{\pm 12.26}$	$78.74_{\pm 6.68}$	$100.00_{\pm 0.00}$
NinaPro	$66.70_{\pm 3.23}$	$65.67_{\pm 3.00}$	$66.31_{\pm 2.98}$	$66.45_{\pm 3.63}$	$66.32_{\pm 2.79}$	$66.17_{\pm 3.30}$	$68.51_{\pm 3.26}$	$68.81_{\pm 3.37}$

For each dataset, Full DTMVC uses the configuration $(\beta^\star, \gamma^\star, \lambda^\star)$ with the highest mean ACC in the three-parameter grid; all four metrics are reported from that same configuration. Controlled ablations should keep the remaining parameters fixed: *w/o Triadic* uses $(0, \gamma^\star, \lambda^\star)$, and *w/o Tensor* uses $(\beta^\star, \gamma^\star, 0)$.

Table 3 shows that Full DTMVC achieves the highest mean ACC on all six datasets and leads in 21 of the 24 dataset–metric comparisons. Compared with Mean-pooling, its ACC is higher by 41.28 and 16.99 percentage points on PASTIS-R and HPC-ODA, respectively. Temporal averaging retains average levels but can discard class-related fluctuations and their temporal arrangement. These observations are consistent with the motivation for retaining temporal information in multi-view clustering.

6 Conclusion

We presented DTMVC for clustering multi-view time series by jointly modeling temporal geometry and shared cross-view structure. Normalized state and first-difference dissimilarities define pairwise and sparse triadic losses, while projection-product consensus and tensor regularization of aligned embeddings couple the views. We developed a two-level proximal MM–iALM algorithm for the resulting nonconvex, nonsmooth problem. Under the stated conditions, the splitting residual vanishes, and every accumulation point of the outer primal iterates corresponds to a first-order stationary point of the original problem. Experiments on six real-world datasets demonstrate the effectiveness of DTMVC and support the joint modeling of temporal information and cross-view structure for clustering. Future work will extend DTMVC to incomplete multi-view time series with missing observations or views.

REFERENCES

Oresti Banos, Rafael Garcia, and Alejandro Saez. MHEALTH. Dataset, 2014. URL https://archive.ics.uci.edu/dataset/319/mhealth+dataset.

Billur Barshan and Kerem Altun. Daily and sports activities. Dataset, 2010. URL https://archive.ics.uci.edu/dataset/256/daily+and+sports+activities.

Nicolas Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023. doi: 10.1017/9781009166164. URL https://www.nicolasboumal.net/book/.

Hongmin Cai, Yu Wang, Fei Qi, Zhuoyao Wang, and Yiu-ming Cheung. Multiview tensor spectral clustering via co-regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(10):6795–6808, 2024. doi: 10.1109/TPAMI.2024.3386828.

Marco Cuturi and Mathieu Blondel. Soft-DTW: A differentiable loss function for time-series. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 894–903, 2017.

Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pp. 2352–2359, 2021. doi: 10.24963/ijcai.2021/324.

Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. Unsupervised scalable representation learning for multivariate time series. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Hongchang Gao, Feiping Nie, Xuelong Li, and Heng Huang. Multi-view subspace clustering. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4238–4246, 2015.

Abhishek Kumar, Piyush Rai, and Hal Daumé, III. Co-regularized multi-view spectral clustering. In *Advances in Neural Information Processing Systems*, volume 24, 2011.

Yurui Li, Mingjing Du, Xiang Jiang, and Nan Zhang. Contrastive learning-based multi-view clustering for incomplete multivariate time series. *Information Fusion*, 117:102812, 2025. doi: 10.1016/j.inffus.2024.102812.

Zhiyu Liang, Jianfeng Zhang, Chen Liang, Hongzhi Wang, Zheng Liang, and Lujia Pan. A shapelet-based framework for unsupervised multivariate time series representation learning. *Proceedings of the VLDB Endowment*, 17(3):386–399, 2023. doi: 10.14778/3632093.3632103.

Bruno Monteiro Rocha Lima, Thiago Eustaquio Alves de Oliveira, and Vinicius Prado da Fonseca. Multimodal tactile texture dataset. Dataset, version 1, 2023. URL https://data.mendeley.com/datasets/n666tk4mw9/1.

Jiyuan Liu, Xinwang Liu, Chuankun Li, Xinhang Wan, Hao Tan, Yi Zhang, Weixuan Liang, Qian Qu, Yu Feng, Renxiang Guan, and Ke Liang. Large-scale multi-view tensor clustering with implicit linear kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20727–20736, 2025. doi: 10.1109/CVPR52734.2025.01930.

Zhen Long, Qiyuan Wang, Yazhou Ren, Yipeng Liu, and Ce Zhu. S^2MVTC: A simple yet efficient scalable multi-view tensor clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26213–26222, 2024. doi: 10.1109/CVPR52733.2024.02274.

Qianli Ma, Jiawei Zheng, Sen Li, and Garrison W. Cottrell. Learning representations for time series clustering. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Alessio Netti, Daniele Tafani, Michael Ott, and Martin Schulz. Correlation-wise smoothing: Lightweight knowledge extraction for HPC monitoring data. In *2021 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pp. 2–12. IEEE, 2021. doi: 10.1109/IPDPS49936.2021.00010.

John Paparrizos and Luis Gravano. k-Shape: Efficient and accurate clustering of time series. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pp. 1855–1870, 2015. doi: 10.1145/2723372.2737793.

Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3): 127–239, 2014. doi: 10.1561/2400000003. URL https://web.stanford.edu/~boyd/papers/prox_algs.html.

Stefano Pizzolato, Luca Tagliapietra, Matteo Cognolato, Monica Reggiani, Henning Müller, and Manfredo Atzori. Comparison of six electromyography acquisition setups on hand movement classification tasks. *PLOS ONE*, 12(10):e0186132, 2017. doi: 10.1371/journal.pone.0186132.

Vivien Sainte Fare Garnot and Loic Landrieu. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

Vivien Sainte Fare Garnot, Loic Landrieu, and Nesrine Chehata. Multi-modal temporal attention models for crop mapping from satellite time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022. doi: 10.1016/j.isprsjprs.2022.03.012.

Renat Sergazinov, Armeen Taeb, and Irina Gaynanova. A spectral method for multi-view subspace learning using the product of projections, 2025. URL https://arxiv.org/abs/2410.19125v2. arXiv:2410.19125v2, 13 August 2025.

Zexi Tan, Xiaopeng Luo, Yunlin Liu, and Yiqun Zhang. Mask the redundancy: Evolving masking representation learning for multivariate time-series clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(30):25787–25795, 2026. doi: 10.1609/aaai.v40i30.39777.

Sana Tonekaboni, Danny Eytan, and Anna Goldenberg. Unsupervised representation learning for time series with temporal neighborhood coding. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=8qDwejCuCN.

Jianlong Wu, Zhouchen Lin, and Hongbin Zha. Essential tensor learning for multi-view spectral clustering. *IEEE Transactions on Image Processing*, 28(12):5910–5922, 2019. doi: 10.1109/TIP.2019.2916740.

Jianlong Wu, Xingyu Xie, Liqiang Nie, Zhouchen Lin, and Hongbin Zha. Unified graph and low-rank tensor learning for multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 6388–6395, 2020. doi: 10.1609/aaai.v34i04.6109.

Rongkai Xia, Yan Pan, Lei Du, and Jian Yin. Robust multi-view spectral clustering via low-rank and sparse decomposition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, 2014.

Yuan Xie, Dacheng Tao, Wensheng Zhang, Yan Liu, Lei Zhang, and Yanyun Qu. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *International Journal of Computer Vision*, 126(11):1157–1179, 2018. doi: 10.1007/s11263-018-1086-2.

Deng Xu, Chao Zhang, Zechao Li, Chunlin Chen, and Huaxiong Li. Fast disentangled slim tensor learning for multi-view clustering. *IEEE Transactions on Multimedia*, 27:1254–1265, 2025. doi: 10.1109/TMM.2024.3521754.

Shuochao Yao, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek Abdelzaher. DeepSense: A unified deep learning framework for time-series mobile sensing data processing. In *Proceedings of the 26th International Conference on World Wide Web*, pp. 351–360, 2017.

Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. TS2Vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 8980–8987, 2022. doi: 10.1609/aaai.v36i8.20881.

George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. A transformer-based framework for multivariate time series representation learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 2114–2124, 2021. doi: 10.1145/3447548.3467401.

Changqing Zhang, Huazhu Fu, Si Liu, Guangcan Liu, and Xiaochun Cao. Low-rank tensor constrained multiview subspace clustering. In *Proceedings of the IEEE International Conference on Computer Vision*, 2015.

Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in Neural Information Processing Systems*, volume 19, 2006.

A Detailed Optimization

A.1 Split Formulation

For each view, let $\Delta_v \in \mathbb{R}^{n\times n}$ be the fixed symmetric dissimilarity matrix and let $\mathcal{W}_v \in \mathbb{R}^{n\times n\times n}$ be the fixed, normalized supersymmetric tensor constructed from the fixed sparse triplet set $\mathcal{T}_v$ defined below. Write $F_v = [f_{v,1}, \dots, f_{v,c}]$ and define

$$\begin{aligned} \Phi_{3,v}(F_v) &= \sum_{q=1}^{c}\sum_{i,j,k=1}^{n} (\mathcal{W}_v)_{ijk}(F_v)_{iq}(F_v)_{jq}(F_v)_{kq}, \\ E_v(F_v) &= \tfrac{1}{2}\operatorname{Tr}(F_v^\top \Delta_v F_v) - b_{2,v} + \beta\big(\sqrt{c}\,\|\mathcal{W}_v\|_F - \Phi_{3,v}(F_v)\big). \end{aligned} \tag{A.1}$$

Here $b_{2,v} = \frac{1}{2}\sum_{j=1}^{c}\lambda_j^{\uparrow}(\Delta_v)$, where $\lambda_j^{\uparrow}$ denotes the jth smallest eigenvalue. The two shifted losses are nonnegative on $F_v^\top F_v = I_c$: the quadratic minimum is $b_{2,v}$, and $\|\sum_q f_{v,q}^{\otimes 3}\|_F = \sqrt{c}$ gives $|\Phi_{3,v}(F_v)| \le \sqrt{c}\,\|\mathcal{W}_v\|_F$. These shifts must be retained when updating the view weights.

Let $P_v = F_v F_v^\top$, $P_H = HH^\top$, and $N_p = V(V-1)/2$, where $V \ge 2$. Define

$$M_\times = \frac{1}{2N_p}\sum_{u<v}(P_u P_v + P_v P_u), \qquad \widetilde{\mathcal{F}}(:,:,v) = F_v R_v. \tag{A.2}$$

The constraints are $F_v, H \in \mathrm{St}(n,c)$, $R_v \in \mathrm{O}(c)$, and $\boldsymbol{\alpha} \in \Delta_V$, where $\mathrm{St}(p,c) = \{X \in \mathbb{R}^{p\times c} : X^\top X = I_c\}$, $\mathrm{O}(c) = \mathrm{St}(c,c)$, and $\Delta_V = \{\boldsymbol{a} \in \mathbb{R}^V : a_v \ge 0,\ \sum_v a_v = 1\}$. The alignment matrices R_v enter only through $\widetilde{\mathcal{F}}$.

Introduce $\mathcal{J} \in \mathbb{R}^{n\times c\times V}$ and write $\mathbf{F} = (F_1, \dots, F_V)$, $\mathbf{R} = (R_1, \dots, R_V)$, and $\Theta = (\mathbf{F}, \mathbf{R}, H, \boldsymbol{\alpha}, \mathcal{J})$. The split problem is

$$\begin{aligned} \min_{\Theta}\quad & \mathcal{Q}(\Theta) = \sum_{v=1}^{V}\alpha_v^r E_v(F_v) - \gamma\operatorname{Tr}(H^\top M_\times H) + \lambda\|\mathcal{J}\|_{\mathrm{TNN}} \\ & \text{s.t.}\quad F_v, H \in \mathrm{St}(n,c),\quad R_v \in \mathrm{O}(c),\quad \boldsymbol{\alpha} \in \Delta_V,\quad \widetilde{\mathcal{F}} = \mathcal{J}. \end{aligned} \tag{A.3}$$

Let $r > 1$ and $\beta, \gamma, \lambda \ge 0$. Write $\mathcal{R}(\Theta) = \widetilde{\mathcal{F}} - \mathcal{J}$ for the split residual. For multiplier $\mathcal{Y}$ and $\rho > 0$, the augmented Lagrangian is

$$\begin{aligned} L_\rho(\Theta; \mathcal{Y}) &= \mathcal{Q}(\Theta) + \langle \mathcal{Y}, \mathcal{R}(\Theta)\rangle + \tfrac{\rho}{2}\|\mathcal{R}(\Theta)\|_F^2 \\ &= \mathcal{Q}(\Theta) + \tfrac{\rho}{2}\|\widetilde{\mathcal{F}} - \mathcal{J} + \mathcal{U}\|_F^2 - \tfrac{\rho}{2}\|\mathcal{U}\|_F^2, \qquad \mathcal{U} = \mathcal{Y}/\rho. \end{aligned} \tag{A.4}$$

A.2 Primal Block Updates

Fix outer iteration s. Throughout its inner loop, $\mathcal{Y}^s$, ρ_s, and $\mathcal{U}^s = \mathcal{Y}^s/\rho_s$ remain fixed. Write $\Theta^t = \Theta^{s,t}$, where t indexes complete inner sweeps. Each sweep updates $\{F_v\}_{v=1}^V$, $\{R_v\}_{v=1}^V$, H, $\boldsymbol{\alpha}$, and $\mathcal{J}$, in that order. For a thin SVD $A = U\Sigma V^\top$ with c orthonormal columns, write $\mathrm{Polar}(A) = UV^\top$; when A is rank deficient, complete the singular vectors to c columns.

Updating F_v. The view embeddings are updated sequentially. Set $S_{-v}^t = \sum_{u<v} P_u^{t+1} + \sum_{u>v} P_u^t$, $C_v^t = P_H^t S_{-v}^t + S_{-v}^t P_H^t$, and $B_v^t = \mathcal{J}_v^t - \mathcal{U}_v^s$, where tensor subscripts denote frontal slices. The F_v subproblem is $\min_{F^\top F = I_c} g_v^t(F)$, where

$$g_v^t(F) = (\alpha_v^t)^r E_v(F) - \frac{\gamma}{2N_p}\operatorname{Tr}(F^\top C_v^t F) + \frac{\rho_s}{2}\|FR_v^t - B_v^t\|_F^2. \tag{A.5}$$

The consensus term follows by collecting every pair containing view v in (A.2) and using cyclicity of the trace. Since R_v^t is orthogonal, $\|FR_v^t\|_F^2 = \|F\|_F^2$, which equals c on the feasible set. Subtracting this constant gives

$$\begin{aligned} q_v^t(F) &= g_v^t(F) - \frac{\rho_s}{2}\|FR_v^t\|_F^2 \\ &= (\alpha_v^t)^r E_v(F) - \frac{\gamma}{2N_p}\operatorname{Tr}(F^\top C_v^t F) - \rho_s\langle B_v^t (R_v^t)^\top, F\rangle_F + \frac{\rho_s}{2}\|B_v^t\|_F^2. \end{aligned} \tag{A.6}$$

Supersymmetry of $\mathcal{W}_v$ gives

$$[\nabla E_v(F)]_{iq} = (\Delta_v F)_{iq} - 3\beta \sum_{j,k=1}^{n} (\mathcal{W}_v)_{ijk} F_{jq} F_{kq},$$
$$G_v^t := \nabla q_v^t(F_v^t) = (\alpha_v^t)^r \nabla E_v(F_v^t) - \frac{\gamma}{N_p} C_v^t F_v^t - \rho_s B_v^t (R_v^t)^\top. \tag{A.7}$$

For a column f, the (i,j) entry of the Hessian of the cubic loss is $-6\beta \sum_k (\mathcal{W}_v)_{ijk} f_k$; its spectral norm is at most $6\beta \|\mathcal{W}_v\|_F \|f\|_2$. Along the straight segment between two feasible embeddings, each column has norm at most one, even though the segment need not remain on the Stiefel manifold. Moreover, $\|\Delta_v\|_2 \le \|\Delta_v\|_\infty$ and $\|C_v^t\|_2 \le 2(V-1)$, where $\|\cdot\|_\infty$ is the maximum absolute row-sum norm. Hence a Frobenius-norm Lipschitz bound for the reduced gradient along every such segment is

$$\ell_{F,v}^t = (\alpha_v^t)^r \big(\|\Delta_v\|_\infty + 6\beta \|\mathcal{W}_v\|_F\big) + \frac{4\gamma}{V}, \qquad \tau_v^t = \ell_{F,v}^t + \epsilon_\tau \max\{1, \ell_{F,v}^t\}, \tag{A.8}$$

with $\epsilon_\tau > 0$. The penalty contributes no curvature because its F-dependent term in (A.6) is linear. For every feasible F, Taylor's bound therefore yields

$$q_v^t(F) \le q_v^t(F_v^t) + \langle G_v^t, F - F_v^t \rangle_F + \frac{\tau_v^t}{2} \|F - F_v^t\|_F^2. \tag{A.9}$$

Using $\|F\|_F^2 = c$, minimizing the right-hand side is equivalent to maximizing $\langle \tau_v^t F_v^t - G_v^t, F \rangle_F$. The SVD gives the exact surrogate minimizer

$$F_v^{t+1} = \mathrm{Polar}(\tau_v^t F_v^t - G_v^t). \tag{A.10}$$

This MM update satisfies $q_v^t(F_v^t) - q_v^t(F_v^{t+1}) \ge (\tau_v^t - \ell_{F,v}^t) \|F_v^{t+1} - F_v^t\|_F^2 / 2$.

Updating R_v. After all view embeddings have been updated, the rotation subproblem with proximal parameter $\eta_R > 0$ is

$$\min_{R^\top R = I_c} \quad \frac{\rho_s}{2} \|F_v^{t+1} R - B_v^t\|_F^2 + \frac{\eta_R}{2} \|R - R_v^t\|_F^2. \tag{A.11}$$

Expanding the squared norms and dropping terms constant on $\mathrm{O}(c)$ gives the maximization of $\langle \rho_s (F_v^{t+1})^\top B_v^t + \eta_R R_v^t, R \rangle_F$. Consequently,

$$R_v^{t+1} = \mathrm{Polar}\big(\rho_s (F_v^{t+1})^\top B_v^t + \eta_R R_v^t\big). \tag{A.12}$$

The constraint is orthogonality; no determinant restriction is imposed.

Updating H. Form $M_\times^{t+1}$ from the updated F_v^{t+1}. With projector proximal parameter $\eta_H > 0$, the subproblem is

$$\min_{H^\top H = I_c} \quad -\gamma \,\mathrm{Tr}(H^\top M_\times^{t+1} H) + \frac{\eta_H}{2} \|H H^\top - P_H^t\|_F^2. \tag{A.13}$$

For rank-c orthogonal projectors, $\|H H^\top - P_H^t\|_F^2 = 2c - 2\,\mathrm{Tr}(H^\top P_H^t H)$. The subproblem therefore maximizes $\mathrm{Tr}\{H^\top (\gamma M_\times^{t+1} + \eta_H P_H^t) H\}$. By the variational characterization of eigenvalues,

$$H^{t+1} = \mathrm{TopEig}_c\big(\gamma M_\times^{t+1} + \eta_H P_H^t\big), \tag{A.14}$$

where TopEig_c selects orthonormal eigenvectors for the c largest algebraic eigenvalues. The matrix need not be positive semidefinite. The proximal term controls $H H^\top$ and is unchanged by a rotation of the columns of H.

Updating $\boldsymbol{\alpha}$. Set $E_v^{t+1} = E_v(F_v^{t+1})$ with the constants in (A.1) included. The weight subproblem is

$$\min_{\boldsymbol{\alpha} \in \Delta_V} \sum_{v=1}^{V} E_v^{t+1} \alpha_v^r. \tag{A.15}$$

If all energies are positive, this problem is strictly convex and its minimizer has positive entries. The multiplier equation $rE_v^{t+1}\alpha_v^{r-1} = \nu$ and the simplex constraint give

$$\alpha_v^{t+1} = \frac{(E_v^{t+1})^{-1/(r-1)}}{\sum_{u=1}^{V}(E_u^{t+1})^{-1/(r-1)}}. \tag{A.16}$$

If $\mathcal{Z}^{t+1} = \{v : E_v^{t+1} = 0\}$ is nonempty, the minimum is zero and every simplex vector supported on $\mathcal{Z}^{t+1}$ is optimal. We select

$$\alpha_v^{t+1} = \begin{cases} 1/|\mathcal{Z}^{t+1}|, & v \in \mathcal{Z}^{t+1}, \\ 0, & v \notin \mathcal{Z}^{t+1}. \end{cases} \tag{A.17}$$

Updating $\mathcal{J}$. Form $\widetilde{\mathcal{F}}^{t+1}(:,:,v) = F_v^{t+1}R_v^{t+1}$ and set $\mathcal{D} = \widetilde{\mathcal{F}}^{t+1} + \mathcal{U}^s$. The final block update minimizes $q_{\mathcal{J}}^t$, defined by

$$q_{\mathcal{J}}^t(\mathcal{J}) = \lambda\|\mathcal{J}\|_{\mathrm{TNN}} + \frac{\rho_s}{2}\|\mathcal{J} - \mathcal{D}\|_F^2. \tag{A.18}$$

For any tensor $\mathcal{A}$, let $\widehat{\mathcal{A}}^{(\ell)} = \sum_{v=1}^{V}\mathcal{A}(:,:,v)\exp\{-2\pi\mathrm{i}(v-1)(\ell-1)/V\}$ be its unnormalized Fourier slices along the third mode. The tensor nuclear norm and Parseval identity take the forms

$$\|\mathcal{A}\|_{\mathrm{TNN}} = \frac{1}{V}\sum_{\ell=1}^{V}\|\widehat{\mathcal{A}}^{(\ell)}\|_*, \qquad \|\mathcal{A}\|_F^2 = \frac{1}{V}\sum_{\ell=1}^{V}\|\widehat{\mathcal{A}}^{(\ell)}\|_F^2. \tag{A.19}$$

Thus the objective in (A.18) becomes

$$\frac{1}{V}\sum_{\ell=1}^{V}\left[\lambda\|\widehat{\mathcal{J}}^{(\ell)}\|_* + \frac{\rho_s}{2}\|\widehat{\mathcal{J}}^{(\ell)} - \widehat{\mathcal{D}}^{(\ell)}\|_F^2\right]. \tag{A.20}$$

For a complex SVD $\widehat{\mathcal{D}}^{(\ell)} = U_\ell\,\mathrm{diag}(\sigma_{\ell,j})V_\ell^*$, each slice is obtained by singular-value thresholding:

$$\begin{aligned} \widehat{\mathcal{J}}^{\,t+1,(\ell)} &= U_\ell\,\mathrm{diag}\big((\sigma_{\ell,j} - \lambda/\rho_s)_+\big)V_\ell^*, \\ \mathcal{J}^{t+1} &= \mathrm{ifft}(\widehat{\mathcal{J}}^{\,t+1}, [], 3). \end{aligned} \tag{A.21}$$

Here $(x)_+ = \max\{x, 0\}$ and * denotes conjugate transpose. The common factor $1/V$ in (A.20) leaves the threshold at λ/ρ_s. For real $\mathcal{D}$, the slices are conjugate symmetric; thresholding preserves this symmetry, so the inverse transform is real.

A.3 Two-Level Stopping Rules and KKT Conditions

Stationarity after a complete sweep. Restore the outer index and form

$$\widehat{\mathcal{Y}}^{s,t+1} = \mathcal{Y}^s + \rho_s\big(\widetilde{\mathcal{F}}^{s,t+1} - \mathcal{J}^{s,t+1}\big). \tag{A.22}$$

This candidate is used to evaluate stationarity; it does not replace $\mathcal{Y}^s$ inside the current inner loop. Consider the ordinary Lagrangian $L(\Theta, \mathcal{Y}) = \mathcal{Q}(\Theta) + \langle \mathcal{Y}, \widetilde{\mathcal{F}} - \mathcal{J}\rangle$. At the completed primal point $\Theta^{s,t+1}$, abbreviate the candidate multiplier by $\widehat{\mathcal{Y}}$ and set $S_{-v} = \sum_{u\neq v} P_u$. For scalar functions, $D_X L = \nabla_X L$ denotes the Euclidean block gradient. Thus

$$\begin{aligned} D_{F_v}L &= \alpha_v^r \nabla E_v(F_v) - \frac{\gamma}{N_p}(P_H S_{-v} + S_{-v}P_H)F_v + \widehat{\mathcal{Y}}_v R_v^\top, \\ D_{R_v}L &= F_v^\top \widehat{\mathcal{Y}}_v, \qquad D_H L = -2\gamma M_\times H. \end{aligned} \tag{A.23}$$

All quantities are evaluated at the completed sweep, so S_{-v} contains every updated projector. With $\widehat{\mathcal{Y}}$ held fixed, these derivatives equal those of $L_{\rho_s}(\Theta; \mathcal{Y}^s)$ at that point.

For an orthonormal matrix X, the tangent projection is $\Pi_X(Z) = Z - X\,\mathrm{sym}(X^\top Z)$, with $\mathrm{sym}(A) = (A + A^\top)/2$. Hence $\mathrm{grad}_X L = \Pi_X(D_X L)$, and the manifold stationarity residual is

$$\mathcal{S}_s^{t+1} = \max\left\{\max_v \|\Pi_{F_v}(D_{F_v}L)\|_F,\ \max_v \|\Pi_{R_v}(D_{R_v}L)\|_F,\ \|\Pi_H(D_H L)\|_F\right\}. \tag{A.24}$$

This residual measures stationarity of (A.3), without the MM surrogate or proximal terms.

The weight and tensor conditions already hold exactly. The weight gradient is $D_{\boldsymbol{\alpha}}\mathcal{Q} = (r\alpha_v^{r-1}E_v(F_v))_{v=1}^V$. Writing N_{Δ_V} for the convex normal cone to the simplex, the weight solution gives $0 \in D_{\boldsymbol{\alpha}}\mathcal{Q} + N_{\Delta_V}(\boldsymbol{\alpha})$. The last tensor update gives

$$0 \in \lambda\partial\|\mathcal{J}^{s,t+1}\|_{\mathrm{TNN}} - \mathcal{Y}^s + \rho_s\big(\mathcal{J}^{s,t+1} - \widetilde{\mathcal{F}}^{s,t+1}\big), \qquad \widehat{\mathcal{Y}}^{s,t+1} \in \lambda\partial\|\mathcal{J}^{s,t+1}\|_{\mathrm{TNN}}. \tag{A.25}$$

Updating $\mathcal{J}$ does not change the energies or weights, so both conditions hold at the point used in (A.24).

Initialization and inner acceptance. Choose feasible $F_v^0, R_v^0, H^0, \boldsymbol{\alpha}^0$, set $\mathcal{J}^0 = \widetilde{\mathcal{F}}^0$ and $\mathcal{Y}^0 = 0$, and take $\rho_0, \eta_R, \eta_H, \epsilon_\tau > 0$. At outer iteration s, set $\Theta^{s,0} = \Theta^s$. With $\varepsilon_s = \varepsilon_0\xi^s$, $\varepsilon_0 > 0$, and $0 < \xi < 1$, accept the first completed sweep $t_s \geq 1$ for which $\mathcal{S}_s^{t_s} \leq \varepsilon_s$. Until this condition holds, continue the inner updates with the same $\mathcal{Y}^s$ and ρ_s.

Multiplier and penalty updates. Set

$$\Theta^{s+1} = \Theta^{s,t_s}, \qquad \mathcal{Y}^{s+1} = \widehat{\mathcal{Y}}^{s,t_s}, \qquad r_{s+1} = \|\widetilde{\mathcal{F}}^{s+1} - \mathcal{J}^{s+1}\|_F. \tag{A.26}$$

Set $\rho_1 = \rho_0$. For $s \geq 1$, use

$$\rho_{s+1} = \begin{cases} \rho_s, & r_{s+1} \leq \theta r_s, \\ \kappa\rho_s, & r_{s+1} > \theta r_s, \end{cases} \qquad 0 < \theta < 1, \quad \kappa > 1. \tag{A.27}$$

After choosing ρ_{s+1}, recompute $\mathcal{U}^{s+1} = \mathcal{Y}^{s+1}/\rho_{s+1}$. The outer loop stops when

$$r_{s+1} \leq \varepsilon_{\mathrm{feas}}, \qquad \mathcal{S}_s^{t_s} \leq \varepsilon_{\mathrm{stat}}, \tag{A.28}$$

for prescribed positive tolerances.

Relation to the KKT system. At a point satisfying the orthogonality and simplex constraints, the split KKT conditions are

$$\begin{aligned} &\widetilde{\mathcal{F}} = \mathcal{J}, \qquad \Pi_{F_v}(D_{F_v}L) = 0, \quad \Pi_{R_v}(D_{R_v}L) = 0 \quad (\forall v), \\ &\Pi_H(D_H L) = 0, \qquad 0 \in D_{\boldsymbol{\alpha}}\mathcal{Q} + N_{\Delta_V}(\boldsymbol{\alpha}), \qquad \mathcal{Y} \in \lambda\partial\|\mathcal{J}\|_{\mathrm{TNN}}. \end{aligned} \tag{A.29}$$

Let $\mathcal{X} = \{F_1, \ldots, F_V, R_1, \ldots, R_V, H\}$. Define

$$\begin{aligned} \mathcal{K}(\Theta, \mathcal{Y}) = \|\widetilde{\mathcal{F}} - \mathcal{J}\|_F + \max_{X \in \mathcal{X}} \|\Pi_X(D_X L)\|_F \\ + \operatorname{dist}\big(0, D_{\boldsymbol{\alpha}}\mathcal{Q} + N_{\Delta_V}(\boldsymbol{\alpha})\big) + \operatorname{dist}\big(0, \lambda\partial\|\mathcal{J}\|_{\mathrm{TNN}} - \mathcal{Y}\big). \end{aligned} \tag{A.30}$$

At a limit point, $\mathcal{K}(\Theta^\star, \mathcal{Y}^\star)$ is the residual $\mathcal{K}^\star$ in the proof of Theorem 1. At every accepted outer point, the weight and tensor terms vanish exactly, giving $\mathcal{K}(\Theta^{s+1}, \mathcal{Y}^{s+1}) = r_{s+1} + \mathcal{S}_s^{t_s}$. The stopping rule (A.28) therefore bounds the total residual by $\varepsilon_{\mathrm{feas}} + \varepsilon_{\mathrm{stat}}$.

A.4 Cluster Assignment

Let H be the final accepted common embedding. Form $\bar{H}$ by setting $\bar{H}_{i,:} = H_{i,:}/\|H_{i,:}\|_2$ for nonzero rows and $\bar{H}_{i,:} = 0$ otherwise. Apply k-means with $k = c$ to the rows of $\bar{H}$ to obtain labels $\widehat{\boldsymbol{y}} \in \{1, \ldots, c\}^n$. The KKT conditions apply to the accepted embedding H, before row normalization.

Fixed sparse triplet tensors. For each view v, let Δ_v be the fixed dissimilarity matrix and let $\mathcal{N}_v(i)$ contain the $k_{\mathrm{nn}} = \min\{10, n-1\}$ nearest samples to i, excluding i itself. Ties are resolved once during preprocessing. Define the set of unordered triples

$$\mathcal{T}_v = \bigcup_{i=1}^{n} \big\{\{i, j, k\} : j, k \in \mathcal{N}_v(i),\ j < k\big\}.$$

The set union removes triples generated by multiple anchors.

For $e = \{i, j, k\} \in \mathcal{T}_v$ with positive side lengths $(\Delta_v)_{ij}$, $(\Delta_v)_{ik}$, and $(\Delta_v)_{jk}$, let ψ_i, ψ_j, ψ_k be the corresponding comparison-triangle angles and set

$$\omega_{v,e} = \frac{3 - \cos\psi_i - \cos\psi_j - \cos\psi_k}{6}.$$

Triples with a zero side have weight zero; positive collinear triples retain the same weight formula. Assign $\omega_{v,e}$ to all six permutations of e in $\mathcal{A}_v \in \mathbb{R}^{n\times n\times n}$, and set all other entries to zero. Then

$$\|\mathcal{A}_v\|_F^2 = 6\sum_{e\in\mathcal{T}_v} \omega_{v,e}^2, \qquad \mathcal{W}_v = \begin{cases} \mathcal{A}_v/\|\mathcal{A}_v\|_F, & \mathcal{A}_v \neq 0, \\ 0, & \mathcal{A}_v = 0. \end{cases}$$

Thus $\mathcal{W}_v$ is a fixed symmetric tensor with $\|\mathcal{W}_v\|_F \le 1$.

For $F_v = [f_{v,1}, \dots, f_{v,c}]$ satisfying $F_v^\top F_v = I_c$, define

$$\Phi_{3,v}(F_v) = \left\langle \mathcal{W}_v, \sum_{\ell=1}^{c} f_{v,\ell}^{\otimes 3} \right\rangle, \qquad \mathcal{L}_{3,v}(F_v) = \sqrt{c}\,\|\mathcal{W}_v\|_F - \Phi_{3,v}(F_v).$$

Orthogonality gives

$$\left\| \sum_{\ell=1}^{c} f_{v,\ell}^{\otimes 3} \right\|_F^2 = \sum_{\ell,m=1}^{c} (f_{v,\ell}^\top f_{v,m})^3 = c.$$

Consequently, Cauchy–Schwarz yields

$$|\Phi_{3,v}(F_v)| \le \sqrt{c}\,\|\mathcal{W}_v\|_F, \qquad \mathcal{L}_{3,v}(F_v) \ge 0.$$

These statements also hold when $\mathcal{W}_v = 0$. All objectives, view energies, and derivatives below use these fixed sparse tensors. The supports $\mathcal{T}_v$ are not updated during optimization.

B Proof of Theorem 1

Proof. By Lemma 1, every fixed-s inner loop terminates after finitely many complete sweeps. Consider an infinite sequence of accepted outer iterates, with $\varepsilon_s = \varepsilon_0 \xi^s \to 0$, where $\varepsilon_0 > 0$ and $0 < \xi < 1$. Lemma 3 gives boundedness of the accepted primal–dual sequence. Let $\Theta^\star$ be an arbitrary accumulation point of the primal sequence and choose $s_j \to \infty$ such that $\Theta^{s_j+1} \to \Theta^\star$. By boundedness of the multipliers, we may pass to a further subsequence, retaining the same notation and primal limit, such that

$$(\Theta^{s_j+1}, \mathcal{Y}^{s_j+1}) \longrightarrow (\Theta^\star, \mathcal{Y}^\star).$$

Closedness of the constraint sets gives

$$F_v^\star, H^\star \in \mathrm{St}(n, c), \qquad R_v^\star \in \mathrm{O}(c), \qquad \boldsymbol{\alpha}^\star \in \Delta_V.$$

Write $L(\Theta, \mathcal{Y}) = \mathcal{Q}(\Theta) + \langle \mathcal{Y}, \widetilde{\mathcal{F}} - \mathcal{J} \rangle$ for the ordinary Lagrangian, with the multiplier held fixed in all block derivatives. Let $\mathcal{X} = \{F_1, \dots, F_V, R_1, \dots, R_V, H\}$ denote the manifold blocks. A superscript star on a block derivative indicates evaluation at $\Theta^\star$, with multiplier $\mathcal{Y}^\star$ for L. To verify the KKT conditions at $(\Theta^\star, \mathcal{Y}^\star)$, define the total residual

$$\begin{aligned} \mathcal{K}^\star := \|\widetilde{\mathcal{F}}^\star - \mathcal{J}^\star\|_F &+ \max_{X\in\mathcal{X}} \|\Pi_{X^\star}(D_X L^\star)\|_F \\ &+ \mathrm{dist}(0, D_{\boldsymbol{\alpha}}\mathcal{Q}^\star + N_{\Delta_V}(\boldsymbol{\alpha}^\star)) \\ &+ \mathrm{dist}(0, \lambda\partial\|\mathcal{J}^\star\|_{\mathrm{TNN}} - \mathcal{Y}^\star) \\ &=: \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}. \end{aligned} \tag{T1.1}$$

The four terms correspond to split feasibility, manifold stationarity, weight stationarity, and tensor stationarity, respectively. We show that each term is zero.

The term I. Lemma 4 gives

$$\|\widetilde{\mathcal{F}}^{s+1} - \mathcal{J}^{s+1}\|_F \longrightarrow 0. \tag{T1.2}$$

Since $\widetilde{\mathcal{F}}(:,:,v) = F_v R_v$, the aligned tensor depends continuously on the primal variables. Passing to the limit along s_j in (T1.2) yields

$$\widetilde{\mathcal{F}}^\star = \mathcal{J}^\star, \qquad \mathrm{I} = 0. \tag{T1.3}$$

The term II. For each $X \in \mathcal{X}$, Lemma 2 gives

$$\|\Pi_{X^{s+1}}\big(D_X L(\Theta^{s+1}, \mathcal{Y}^{s+1})\big)\|_F \le \varepsilon_s.$$

The block gradients are continuous in $(\Theta, \mathcal{Y})$: the fixed view energies and the consensus term are smooth in the matrix blocks, and the coefficients α_v^r are continuous. The tangent projection $\Pi_X(Z) = Z - X\,\mathrm{sym}(X^\top Z)$ is also continuous in (X, Z). Thus, for a fixed block X,

$$\begin{aligned}\|\Pi_{X^\star}(D_X L^\star)\|_F &\le \big\|\Pi_{X^\star}(D_X L^\star) - \Pi_{X^{s_j+1}}\big(D_X L(\Theta^{s_j+1}, \mathcal{Y}^{s_j+1})\big)\big\|_F \\ &\quad + \big\|\Pi_{X^{s_j+1}}\big(D_X L(\Theta^{s_j+1}, \mathcal{Y}^{s_j+1})\big)\big\|_F \\ &=: \mathrm{II}_1 + \mathrm{II}_2.\end{aligned}$$

The first term tends to zero by continuity, while $\mathrm{II}_2 \le \varepsilon_{s_j} \to 0$. There are finitely many blocks, so $\mathrm{II} = 0$. Using the projection–normal-space identity from Lemma 2, together with $D_{R_v}\mathcal{Q} = 0$, we obtain

$$\begin{aligned}&0 \in D_{F_v}\mathcal{Q}^\star + \mathcal{Y}_v^\star (R_v^\star)^\top + N_{\mathrm{St}(n,c)}(F_v^\star), \quad v = 1, \dots, V, \\ &0 \in (F_v^\star)^\top \mathcal{Y}_v^\star + N_{\mathrm{O}(c)}(R_v^\star), \qquad\qquad\qquad v = 1, \dots, V, \\ &0 \in D_H \mathcal{Q}^\star + N_{\mathrm{St}(n,c)}(H^\star).\end{aligned} \tag{T1.4}$$

Here $\mathcal{Y}_v^\star = \mathcal{Y}^\star(:,:,v)$.

The term III. The weight gradient is

$$D_{\boldsymbol{\alpha}}\mathcal{Q}(\Theta) = \big(r\alpha_v^{r-1} E_v(F_v)\big)_{v=1}^V.$$

Because $r > 1$, this mapping is continuous on the feasible set, including at zero weights. By Lemma 2, $-D_{\boldsymbol{\alpha}}\mathcal{Q}(\Theta^{s+1}) \in N_{\Delta_V}(\boldsymbol{\alpha}^{s+1})$. Equivalently, for every $\boldsymbol{a} \in \Delta_V$,

$$\langle D_{\boldsymbol{\alpha}}\mathcal{Q}(\Theta^{s_j+1}), \boldsymbol{a} - \boldsymbol{\alpha}^{s_j+1}\rangle \ge 0.$$

Passing to the limit gives $\langle D_{\boldsymbol{\alpha}}\mathcal{Q}^\star, \boldsymbol{a} - \boldsymbol{\alpha}^\star\rangle \ge 0$ for every $\boldsymbol{a} \in \Delta_V$. By the definition of the simplex normal cone,

$$0 \in D_{\boldsymbol{\alpha}}\mathcal{Q}^\star + N_{\Delta_V}(\boldsymbol{\alpha}^\star), \qquad \mathrm{III} = 0. \tag{T1.5}$$

The term IV. Lemma 2 gives $\mathcal{Y}^{s+1} \in \lambda\partial\|\mathcal{J}^{s+1}\|_{\mathrm{TNN}}$. The convex subgradient inequality therefore implies, for every conformable tensor $\mathcal{B}$,

$$\lambda\|\mathcal{B}\|_{\mathrm{TNN}} \ge \lambda\|\mathcal{J}^{s_j+1}\|_{\mathrm{TNN}} + \langle \mathcal{Y}^{s_j+1}, \mathcal{B} - \mathcal{J}^{s_j+1}\rangle.$$

Continuity of the TNN and the inner product allows passage to the limit, yielding

$$\lambda\|\mathcal{B}\|_{\mathrm{TNN}} \ge \lambda\|\mathcal{J}^\star\|_{\mathrm{TNN}} + \langle \mathcal{Y}^\star, \mathcal{B} - \mathcal{J}^\star\rangle.$$

Since this holds for every $\mathcal{B}$,

$$0 \in \lambda\partial\|\mathcal{J}^\star\|_{\mathrm{TNN}} - \mathcal{Y}^\star, \qquad \mathrm{IV} = 0. \tag{T1.6}$$

Thus $\mathcal{K}^\star = 0$. Together with the constraint membership above, (T1.3), (T1.4), (T1.5), and (T1.6) establish the KKT conditions of the split formulation.

Stationarity of the original formulation. Write $\mathbf{F} = (F_1, \dots, F_V)$ and $\mathbf{R} = (R_1, \dots, R_V)$. Regard the aligned tensor as a function $\widetilde{\mathcal{F}}(\mathbf{F}, \mathbf{R})$ with frontal slices $F_v R_v$, and set

$$h(\mathbf{F}, \mathbf{R}) = \lambda\|\widetilde{\mathcal{F}}(\mathbf{F}, \mathbf{R})\|_{\mathrm{TNN}}.$$

By (T1.3) and (T1.6), $\mathcal{Y}^\star \in \lambda\partial\|\widetilde{\mathcal{F}}(\mathbf{F}^\star, \mathbf{R}^\star)\|_{\mathrm{TNN}}$. For perturbations $\delta\mathbf{F}, \delta\mathbf{R}$, the exact product expansion is

$$\begin{aligned}&(F_v^\star + \delta F_v)(R_v^\star + \delta R_v) - F_v^\star R_v^\star \\ &\qquad = \delta F_v R_v^\star + F_v^\star \delta R_v + \delta F_v \delta R_v.\end{aligned}$$

Write $\delta = (\delta\mathbf{F}, \delta\mathbf{R})$, with $\|\delta\|^2 = \sum_v (\|\delta F_v\|_F^2 + \|\delta R_v\|_F^2)$. The convex subgradient inequality gives

$$\begin{aligned} &h(\mathbf{F}^\star + \delta\mathbf{F}, \mathbf{R}^\star + \delta\mathbf{R}) - h(\mathbf{F}^\star, \mathbf{R}^\star) \\ &\geq \sum_v \Big[\langle \mathcal{Y}_v^\star (R_v^\star)^\top, \delta F_v \rangle + \langle (F_v^\star)^\top \mathcal{Y}_v^\star, \delta R_v \rangle \Big] - \frac{\|\mathcal{Y}^\star\|_F}{2} \|\delta\|^2, \end{aligned}$$

since

$$\left| \sum_v \langle \mathcal{Y}_v^\star, \delta F_v \delta R_v \rangle \right| \leq \|\mathcal{Y}^\star\|_F \sum_v \|\delta F_v\|_F \|\delta R_v\|_F \leq \frac{\|\mathcal{Y}^\star\|_F}{2} \|\delta\|^2.$$

The coefficient vector

$$\Big(\{\mathcal{Y}_v^\star (R_v^\star)^\top\}_{v=1}^V, \{(F_v^\star)^\top \mathcal{Y}_v^\star\}_{v=1}^V \Big)$$

therefore belongs to the Fréchet subdifferential of h at $(\mathbf{F}^\star, \mathbf{R}^\star)$, and hence to its limiting subdifferential. The first two lines of (T1.4) then imply

$$0 \in (D_{\mathbf{F}} \mathcal{Q}^\star, 0_{\mathbf{R}}) + \partial h(\mathbf{F}^\star, \mathbf{R}^\star) + N_{\mathrm{St}(n,c)^V \times \mathrm{O}(c)^V}(\mathbf{F}^\star, \mathbf{R}^\star). \tag{T1.7}$$

Together with the H-condition in (T1.4) and (T1.5), this establishes first-order stationarity of the original formulation. Since $\Theta^\star$ was arbitrary, every accumulation point of the accepted primal sequence satisfies the claimed KKT conditions. □

Lemma 1. *Fix an outer iteration s and the corresponding pair $(\mathcal{Y}^s, \rho_s)$, with the dissimilarity matrices Δ_v and the sparse tensors $\mathcal{W}_v$ defined above held fixed throughout optimization. Let $r > 1$, $\beta, \gamma, \lambda \geq 0$, and $\rho_s, \eta_R, \eta_H, \epsilon_\tau > 0$. Suppose that the initial blocks satisfy the Stiefel, orthogonal, and simplex constraints, that $\mathcal{J}^0$ is finite, and that each prescribed block update, including the weight update when some view energies vanish, is computed exactly in the stated order. If the inner sequence is continued indefinitely, then*

$$\mathcal{S}_s^{t+1} \longrightarrow 0.$$

Consequently, for every fixed $\varepsilon_s > 0$, the stopping condition $\mathcal{S}_s^{t+1} \leq \varepsilon_s$ is reached after finitely many complete sweeps.

Proof. Fix s and write $\Theta^t = \Theta^{s,t}$ for the complete primal iterate after sweep t, suppressing s from its components. Write $\mathbf{F}^t = (F_1^t, \ldots, F_V^t)$, $P_v^t = F_v^t (F_v^t)^\top$, $P_H^t = H^t (H^t)^\top$, and $\mathcal{U}^s = \mathcal{Y}^s / \rho_s$. The Euclidean orthogonal projection onto the tangent space at X is $\Pi_X(Z) = Z - X\,\mathrm{sym}(X^\top Z)$, where $\mathrm{sym}(A) = (A + A^\top)/2$ (Boumal, 2023, Sec. 7.3). With $\mathcal{Y}^s$ fixed, write $L_{\rho_s}(\Theta) = L_{\rho_s}(\Theta; \mathcal{Y}^s)$. Thus

$$L_{\rho_s}(\Theta) = \mathcal{Q}(\Theta) + \langle \mathcal{Y}^s, \mathcal{R}(\Theta) \rangle + \frac{\rho_s}{2} \|\mathcal{R}(\Theta)\|_F^2, \qquad \mathcal{R}(\Theta) = \widetilde{\mathcal{F}} - \mathcal{J}.$$

Here $\widetilde{\mathcal{F}}_v = F_v R_v$, and $\mathcal{Q}$ is the original objective.

Step 1. Sufficient decrease. Put $\delta Z^t = Z^{t+1} - Z^t$, using $Z = P_H$ for the consensus projector. Let g_v^t be the prescribed Gauss–Seidel objective for the F_v-update, and write the tensor objective as

$$q_J^t(\mathcal{J}) = \lambda \|\mathcal{J}\|_{\mathrm{TNN}} + \frac{\rho_s}{2} \|\mathcal{J} - \widetilde{\mathcal{F}}^{\,t+1} - \mathcal{U}^s\|_F^2.$$

With $(\mathcal{Y}^s, \rho_s)$ fixed, the updates of $F, R, H, \boldsymbol{\alpha}, \mathcal{J}$, in that order, give the identity

$$\begin{aligned} L_{\rho_s}(\Theta^t) - L_{\rho_s}(\Theta^{t+1}) = &\sum_v \big[g_v^t(F_v^t) - g_v^t(F_v^{t+1}) \big] \\ &+ \frac{\rho_s}{2} \sum_v \Big[\|F_v^{t+1} R_v^t - \mathcal{J}_v^t + \mathcal{U}_v^s\|_F^2 - \|F_v^{t+1} R_v^{t+1} - \mathcal{J}_v^t + \mathcal{U}_v^s\|_F^2 \Big] \\ &+ \gamma \,\mathrm{Tr}\big[M_\times^{t+1} (P_H^{t+1} - P_H^t) \big] \\ &+ \sum_v \big[(\alpha_v^t)^r - (\alpha_v^{t+1})^r \big] E_v(F_v^{t+1}) \\ &+ \big[q_J^t(\mathcal{J}^t) - q_J^t(\mathcal{J}^{t+1}) \big] \\ =: \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} + \mathrm{V}. \end{aligned}$$

Here $M_{\times}^{t+1}$ is formed from the updated embeddings $\{F_v^{t+1}\}_{v=1}^V$. The five terms are the decreases in the same augmented Lagrangian caused by the five groups of updates; proximal penalties are excluded.

Bound for I. Define the reduced objective

$$q_v^t(F) = g_v^t(F) - \frac{\rho_s}{2}\|FR_v^t\|_F^2 = g_v^t(F) - \frac{\rho_s}{2}\|F\|_F^2.$$

The subtracted term equals $\rho_s c/2$ at both feasible endpoints. Hence

$$\begin{aligned}
\mathrm{I} &= \sum_v \big[q_v^t(F_v^t) - q_v^t(F_v^{t+1})\big] \\
&= \sum_v \Big[-\langle \nabla q_v^t(F_v^t), \delta F_v^t\rangle - \frac{\tau_v^t}{2}\|\delta F_v^t\|_F^2\Big] \\
&\quad + \sum_v \Big[q_v^t(F_v^t) - q_v^t(F_v^{t+1}) + \langle \nabla q_v^t(F_v^t), \delta F_v^t\rangle + \frac{\tau_v^t}{2}\|\delta F_v^t\|_F^2\Big] \\
&=: \mathrm{I}_1 + \mathrm{I}_2.
\end{aligned}$$

For I_1, the old point F_v^t is feasible for the quadratic surrogate. Comparison with its exact minimizer F_v^{t+1} gives

$$\langle \nabla q_v^t(F_v^t), \delta F_v^t\rangle + \frac{\tau_v^t}{2}\|\delta F_v^t\|_F^2 \le 0.$$

Summing over v yields $\mathrm{I}_1 \ge 0$.

For I_2, every column along the segment joining two feasible matrices has norm at most one. For matrices A, B with this column norm bound, the identity

$$a\otimes a - b\otimes b = (a-b)\otimes a + b\otimes(a-b)$$

gives

$$\|A\odot A - B\odot B\|_F \le 2\|A-B\|_F,$$

where $\odot$ denotes the columnwise Kronecker product. Since $\mathcal{W}_v$ is fixed and symmetric,

$$\nabla\Phi_{3,v}(F) = 3(\mathcal{W}_v)_{(1)}(F\odot F),$$

where $(\mathcal{W}_v)_{(1)}$ is its mode-one unfolding. Therefore,

$$\begin{aligned}
&\|\nabla\Phi_{3,v}(A) - \nabla\Phi_{3,v}(B)\|_F \\
&\quad\le 3\|(\mathcal{W}_v)_{(1)}\|_2\|A\odot A - B\odot B\|_F \\
&\quad\le 6\|\mathcal{W}_v\|_F\|A-B\|_F.
\end{aligned}$$

Thus the cubic-gradient Lipschitz constant remains bounded by $6\|\mathcal{W}_v\|_F$ for the fixed sparse construction. Symmetry gives $\|\Delta_v\|_2 \le \|\Delta_v\|_\infty$. For $S_{-v}^t = \sum_{u<v} P_u^{t+1} + \sum_{u>v} P_u^t$, the projector norm bounds give

$$\|P_H^t S_{-v}^t + S_{-v}^t P_H^t\|_2 \le 2(V-1), \qquad \frac{2\gamma(V-1)}{\binom{V}{2}} = \frac{4\gamma}{V}.$$

The remaining term in q_v^t is affine in F. Thus the reduced gradient has Lipschitz bound $\ell_{F,v}^t$, with the prescribed curvature

$$\begin{aligned}
\ell_{F,v}^t &= (\alpha_v^t)^r\big(\|\Delta_v\|_\infty + 6\beta\|\mathcal{W}_v\|_F\big) + \frac{4\gamma}{V}, \\
\tau_v^t &= \ell_{F,v}^t + \epsilon_\tau \max\{1, \ell_{F,v}^t\}.
\end{aligned}$$

The descent lemma (Boumal, 2023, Prop. 10.53), applied in the ambient Euclidean space with the gradient bound along the segment joining F_v^t and F_v^{t+1}, gives

$$q_v^t(F_v^{t+1}) - q_v^t(F_v^t) \le \langle \nabla q_v^t(F_v^t), \delta F_v^t\rangle + \frac{\ell_{F,v}^t}{2}\|\delta F_v^t\|_F^2.$$

Substitution into I_2 yields

$$\mathrm{I}_2 \geq \frac{1}{2}\sum_v (\tau_v^t - \ell_{F,v}^t)\|\delta F_v^t\|_F^2 \geq \frac{\epsilon_\tau}{2}\sum_v \|\delta F_v^t\|_F^2.$$

Combining the bounds for I_1 and I_2,

$$\mathrm{I} \geq \frac{\epsilon_\tau}{2}\sum_v \|\delta F_v^t\|_F^2.$$

Bound for II. The R_v-subproblems are separable. Each previous block R_v^t is feasible and has zero proximal penalty. Comparison with the exact proximal minimizers gives

$$\begin{aligned}\frac{\rho_s}{2}\sum_v \|F_v^{t+1}R_v^{t+1} - \mathcal{J}_v^t + \mathcal{U}_v^s\|_F^2 + \frac{\eta_R}{2}\sum_v \|\delta R_v^t\|_F^2 \\ \leq \frac{\rho_s}{2}\sum_v \|F_v^{t+1}R_v^t - \mathcal{J}_v^t + \mathcal{U}_v^s\|_F^2.\end{aligned}$$

Hence $\mathrm{II} \geq \eta_R \sum_v \|\delta R_v^t\|_F^2/2$.

Bound for III. The H-update minimizes its block objective plus $\eta_H\|HH^\top - P_H^t\|_F^2/2$. Comparing its minimizer with H^t gives

$$-\gamma \operatorname{Tr}(M_\times^{t+1}P_H^{t+1}) + \frac{\eta_H}{2}\|\delta P_H^t\|_F^2 \leq -\gamma \operatorname{Tr}(M_\times^{t+1}P_H^t).$$

Thus $\mathrm{III} \geq \eta_H\|\delta P_H^t\|_F^2/2$.

Bound for IV. Since $\boldsymbol{\alpha}^t$ is feasible for the exact weight update,

$$\mathrm{IV} = \sum_v \big[(\alpha_v^t)^r - (\alpha_v^{t+1})^r\big] E_v(F_v^{t+1}) \geq 0.$$

Bound for V. The function q_J^t is ρ_s-strongly convex and $0 \in \partial q_J^t(\mathcal{J}^{t+1})$. Consequently,

$$\mathrm{V} = q_J^t(\mathcal{J}^t) - q_J^t(\mathcal{J}^{t+1}) \geq \frac{\rho_s}{2}\|\delta\mathcal{J}^t\|_F^2.$$

Combining the five bounds yields

$$\begin{aligned}L_{\rho_s}(\Theta^t) - L_{\rho_s}(\Theta^{t+1}) \geq \frac{\epsilon_\tau}{2}\sum_v \|\delta F_v^t\|_F^2 + \frac{\eta_R}{2}\sum_v \|\delta R_v^t\|_F^2 \\ + \frac{\eta_H}{2}\|\delta P_H^t\|_F^2 + \frac{\rho_s}{2}\|\delta\mathcal{J}^t\|_F^2.\end{aligned} \tag{L1.1}$$

As established in the problem restatement, $E_v(F_v) \geq 0$ for every view. Each product of orthogonal projectors has operator norm at most one, so $\|M_\times\|_2 \leq 1$ and $\mathcal{Q}(\Theta) \geq -\gamma c$. Completing the square gives

$$L_{\rho_s}(\Theta) = \mathcal{Q}(\Theta) + \frac{\rho_s}{2}\|\mathcal{R}(\Theta) + \mathcal{U}^s\|_F^2 - \frac{\|\mathcal{Y}^s\|_F^2}{2\rho_s} \geq -\gamma c - \frac{\|\mathcal{Y}^s\|_F^2}{2\rho_s}.$$

Summing (L1.1) from $t = 0$ to $T-1$ bounds the sum of its right-hand side by $L_{\rho_s}(\Theta^0) + \gamma c + \|\mathcal{Y}^s\|_F^2/(2\rho_s)$, independently of T. The four nonnegative terms are therefore summable. Their strictly positive coefficients imply that each norm on the right-hand side tends to zero. Since there are finitely many blocks,

$$d_t := \sum_v \big(\|\delta F_v^t\|_F + \|\delta R_v^t\|_F\big) + \|\delta P_H^t\|_F + \|\delta\mathcal{J}^t\|_F \longrightarrow 0. \tag{L1.2}$$

The tensor proximal map is nonexpansive (Parikh & Boyd, 2014, Sec. 2.3) and maps zero to zero. Hence

$$\|\mathcal{J}^{t+1}\|_F \leq \|\widetilde{\mathcal{F}}^{t+1} + \mathcal{Y}^s/\rho_s\|_F \leq \sqrt{Vc} + \|\mathcal{Y}^s\|_F/\rho_s.$$

Together with finite $\mathcal{J}^0$ and compactness of the Stiefel, orthogonal, and simplex constraints, this proves boundedness of all iterates. The curvature formula and $0 \leq \alpha_v^t \leq 1$ also give $\sup_t \tau_v^t < \infty$.

Throughout the proof, C_s denotes a finite constant independent of t, which may vary from line to line.

Step 2. The F_v-residuals. Write $E(\mathbf{F}) = (E_1(F_1), \dots, E_V(F_V))$, let $a(E(\mathbf{F}))$ be the prescribed exact weight selection, and set

$$\mathcal{G}_v(\mathbf{F}) = a_v(E(\mathbf{F}))^r \Pi_{F_v}\big(\nabla E_v(F_v)\big).$$

For $t \geq 1$, the exact weight update gives $\boldsymbol{\alpha}^t = a(E(\mathbf{F}^t))$. Fix v and $t \geq 1$, and write

$$f = F_v^t, \qquad f_+ = F_v^{t+1}, \qquad e = (\alpha_v^t)^r \nabla E_v(f).$$

Let $S_+ = \sum_{u \neq v} P_u^{t+1}$. Using S_{-v}^t from Step 1, define

$$\mathsf{C} = P_H^t S_{-v}^t + S_{-v}^t P_H^t,$$
$$\mathsf{C}_+ = P_H^{t+1} S_+ + S_+ P_H^{t+1}.$$

The corresponding coupling terms are

$$B = -\frac{\gamma}{N_p}\mathsf{C} f - \rho_s(\mathcal{J}_v^t - \mathcal{U}_v^s)(R_v^t)^\top,$$
$$B_+ = -\frac{\gamma}{N_p}\mathsf{C}_+ f_+ - \rho_s(\mathcal{J}_v^{t+1} - \mathcal{U}_v^s)(R_v^{t+1})^\top,$$

where $N_p = \binom{V}{2}$.

The reduced gradient used in the update is $\nabla q_v^t(f) = e + B$. The exact surrogate minimizer satisfies the first-order necessary condition on the Stiefel manifold (Boumal, 2023, Props. 4.5–4.6):

$$\Pi_{f_+}(e + B + \tau_v^t \delta F_v^t) = 0.$$

At the completed sweep,

$$D_{F_v} L_{\rho_s}(\Theta^{t+1}) = (\alpha_v^{t+1})^r \nabla E_v(f_+) + B_+ + \rho_s f_+.$$

Projecting this identity, subtracting the surrogate optimality condition, and using $\Pi_{f_+}(f_+) = 0$ and $\mathcal{G}_v(\mathbf{F}^t) = \Pi_f(e)$, we obtain

$$\begin{aligned} \Pi_{f_+}\big(D_{F_v} L_{\rho_s}(\Theta^{t+1})\big) &= \big[\mathcal{G}_v(\mathbf{F}^{t+1}) - \mathcal{G}_v(\mathbf{F}^t)\big] \\ &\quad + (\Pi_f - \Pi_{f_+})(e) \\ &\quad + \Pi_{f_+}(B_+ - B - \tau_v^t \delta F_v^t) \\ &=: \mathrm{I} + \mathrm{II} + \mathrm{III}. \end{aligned} \tag{L1.3}$$

Bound for I. We prove continuity of $\mathcal{G}_v$ at an arbitrary point $\mathbf{F}^\star$ of the product Stiefel manifold.

If $E_u(F_u^\star) > 0$ for every u, the energies remain positive in a neighborhood of $\mathbf{F}^\star$. There, the unique weight minimizer is

$$a_v(E) = \frac{E_v^{-1/(r-1)}}{\sum_u E_u^{-1/(r-1)}}.$$

The weight formula and the smoothness of the projected energy gradient imply continuity of $\mathcal{G}_v$ at $\mathbf{F}^\star$.

Otherwise, some view energy vanishes at $\mathbf{F}^\star$. Whenever some energy is zero, the minimum weighted energy is zero. Every minimizing weight vector is therefore supported on the zero-energy views. For such a view, nonnegativity of the smooth function E_v on the Stiefel manifold gives $\Pi_{F_v} \nabla E_v(F_v) = 0$. For a positive-energy view, the minimizing weight is zero. Thus all $\mathcal{G}_v$ vanish at every point with a zero energy.

If $E_v(F_v^\star) = 0$, then, as $\mathbf{F} \to \mathbf{F}^\star$,

$$\|\mathcal{G}_v(\mathbf{F})\|_F \leq \|\Pi_{F_v} \nabla E_v(F_v)\|_F \longrightarrow 0,$$

since $0 \le a_v^r \le 1$ and the projected gradient is continuous and zero at $F_v^\star$. If $E_v(F_v^\star) > 0$, choose u with $E_u(F_u^\star) = 0$. Along points with all energies positive,

$$a_v(E)^r \le \left(\frac{E_u(F_u)}{E_v(F_v)}\right)^{r/(r-1)} \longrightarrow 0.$$

Boundedness of the projected gradient then gives $\mathcal{G}_v(\mathbf{F}) \to 0$. Along the remaining points, $\mathcal{G}_v(\mathbf{F}) = 0$ already. This proves continuity at $\mathbf{F}^\star$.

The product Stiefel manifold is compact, so $\mathcal{G}_v$ is uniformly continuous. By (L1.2), $\sum_u \|F_u^{t+1} - F_u^t\|_F \to 0$. Consequently,

$$\|\mathrm{I}\|_F = \|\mathcal{G}_v(\mathbf{F}^{t+1}) - \mathcal{G}_v(\mathbf{F}^t)\|_F \longrightarrow 0.$$

Bound for II. Using $\Pi_X(Z) = Z - X\,\mathrm{sym}(X^\top Z)$, we have

$$\begin{aligned}\mathrm{II} &= (\Pi_f - \Pi_{f_+})(e)\\ &= \delta F_v^t\,\mathrm{sym}(f^\top e) + f_+\,\mathrm{sym}((\delta F_v^t)^\top e)\\ &=: \mathrm{II}_1 + \mathrm{II}_2.\end{aligned}$$

Using $\|f\|_2 = \|f_+\|_2 = 1$ and $\|\mathrm{sym}(Z)\|_F \le \|Z\|_F$, we obtain

$$\begin{aligned}\|\mathrm{II}_1\|_F &\le \|\delta F_v^t\|_F \|e\|_F,\\ \|\mathrm{II}_2\|_F &\le \|(\delta F_v^t)^\top e\|_F \le \|\delta F_v^t\|_F \|e\|_F.\end{aligned}$$

Since ∇E_v is bounded on the Stiefel manifold and $0 \le \alpha_v^t \le 1$, $\|e\|_F \le C_s$. Therefore

$$\|\mathrm{II}\|_F \le \|\mathrm{II}_1\|_F + \|\mathrm{II}_2\|_F \le C_s \|\delta F_v^t\|_F \longrightarrow 0.$$

Bound for III. Expanding $B_+ - B$,

$$\begin{aligned}\mathrm{III} = &-\frac{\gamma}{N_p}\Pi_{f_+}(\mathsf{C}_+ f_+ - \mathsf{C} f)\\ &- \rho_s \Pi_{f_+}\left[(\mathcal{J}_v^{t+1} - \mathcal{U}_v^s)(R_v^{t+1})^\top - (\mathcal{J}_v^t - \mathcal{U}_v^s)(R_v^t)^\top\right]\\ &- \tau_v^t \Pi_{f_+}(\delta F_v^t)\\ =: &\ \mathrm{III}_1 + \mathrm{III}_2 + \mathrm{III}_3.\end{aligned}$$

For III_1, the identity

$$P_u^{t+1} - P_u^t = \delta F_u^t (F_u^{t+1})^\top + F_u^t (\delta F_u^t)^\top$$

implies $\|S_+ - S_{-v}^t\|_F \le 2\sum_{u>v} \|\delta F_u^t\|_F$. Moreover,

$$\begin{aligned}\mathsf{C}_+ - \mathsf{C} &= \delta P_H^t S_+ + S_+ \delta P_H^t + P_H^t (S_+ - S_{-v}^t) + (S_+ - S_{-v}^t) P_H^t,\\ \mathsf{C}_+ f_+ - \mathsf{C} f &= \mathsf{C}_+ \delta F_v^t + (\mathsf{C}_+ - \mathsf{C}) f.\end{aligned}$$

Since $\|S_+\|_2 \le V - 1$, $\|\mathsf{C}_+\|_2 \le 2(V-1)$, and $\|P_H^t\|_2 = \|f\|_2 = 1$,

$$\begin{aligned}\|\mathsf{C}_+ - \mathsf{C}\|_F &\le 2(V-1)\|\delta P_H^t\|_F + 2\|S_+ - S_{-v}^t\|_F,\\ \|\mathsf{C}_+ f_+ - \mathsf{C} f\|_F &\le 2(V-1)\|\delta F_v^t\|_F + \|\mathsf{C}_+ - \mathsf{C}\|_F \le C_s d_t.\end{aligned}$$

The orthogonal projection Π_{f_+} is nonexpansive, so $\|\mathrm{III}_1\|_F \le C_s d_t$.

For III_2, expand the product difference as

$$\begin{aligned}&(\mathcal{J}_v^{t+1} - \mathcal{U}_v^s)(R_v^{t+1})^\top - (\mathcal{J}_v^t - \mathcal{U}_v^s)(R_v^t)^\top\\ &\qquad = \delta\mathcal{J}_v^t (R_v^{t+1})^\top + (\mathcal{J}_v^t - \mathcal{U}_v^s)(\delta R_v^t)^\top.\end{aligned}$$

Orthogonality of R_v^{t+1}, boundedness from Step 1, and nonexpansiveness of Π_{f_+} give

$$\|\mathrm{III}_2\|_F \le \rho_s \|\delta\mathcal{J}_v^t\|_F + C_s \|\delta R_v^t\|_F \le C_s d_t.$$

For III_3, boundedness of τ_v^t gives

$$\|\mathrm{III}_3\|_F \le \tau_v^t \|\delta F_v^t\|_F \le C_s d_t.$$

Thus $\|\mathrm{III}\|_F \le C_s d_t \to 0$. Combining the bounds in (L1.3),

$$\left\|\Pi_{f_+}\left(D_{F_v} L_{\rho_s}(\Theta^{t+1})\right)\right\|_F \le \|\mathrm{I}\|_F + C_s d_t \longrightarrow 0.$$

Since V is finite, it follows that

$$\max_v \left\|\Pi_{F_v^{t+1}}\left(D_{F_v} L_{\rho_s}(\Theta^{t+1})\right)\right\|_F \longrightarrow 0. \tag{L1.4}$$

Step 3. The remaining residuals.

The R_v-residuals. The R_v-update minimizes

$$\frac{\rho_s}{2}\|F_v^{t+1} R - \mathcal{J}_v^t + \mathcal{U}_v^s\|_F^2 + \frac{\eta_R}{2}\|R - R_v^t\|_F^2 \quad \text{over } R^\top R = I_c.$$

Its first-order condition is

$$\Pi_{R_v^{t+1}}\left[\rho_s (F_v^{t+1})^\top (F_v^{t+1} R_v^{t+1} - \mathcal{J}_v^t + \mathcal{U}_v^s) + \eta_R \delta R_v^t\right] = 0.$$

At the completed sweep,

$$D_{R_v} L_{\rho_s}(\Theta^{t+1}) = \rho_s (F_v^{t+1})^\top (F_v^{t+1} R_v^{t+1} - \mathcal{J}_v^{t+1} + \mathcal{U}_v^s).$$

Subtracting the projected optimality condition gives

$$\begin{aligned}\Pi_{R_v^{t+1}}\left(D_{R_v} L_{\rho_s}(\Theta^{t+1})\right) &= -\eta_R \Pi_{R_v^{t+1}}(\delta R_v^t) - \rho_s \Pi_{R_v^{t+1}}\left((F_v^{t+1})^\top \delta \mathcal{J}_v^t\right) \\ &=: \mathrm{I}_1 + \mathrm{I}_2.\end{aligned}$$

Nonexpansiveness of $\Pi_{R_v^{t+1}}$ and $\|F_v^{t+1}\|_2 = 1$ give

$$\begin{aligned}\|\mathrm{I}_1\|_F &\le \eta_R \|\delta R_v^t\|_F, \\ \|\mathrm{I}_2\|_F &\le \rho_s \|(F_v^{t+1})^\top \delta \mathcal{J}_v^t\|_F \le \rho_s \|\delta \mathcal{J}_v^t\|_F.\end{aligned}$$

The triangle inequality therefore yields

$$\left\|\Pi_{R_v^{t+1}}\left(D_{R_v} L_{\rho_s}(\Theta^{t+1})\right)\right\|_F \le \eta_R \|\delta R_v^t\|_F + \rho_s \|\delta \mathcal{J}_v^t\|_F \le C_s d_t \longrightarrow 0, \tag{L1.5}$$

where $\|\delta \mathcal{J}_v^t\|_F \le \|\delta \mathcal{J}^t\|_F$ and (L1.2) were used.

The H-residual. The first-order condition for the H-update is

$$\Pi_{H^{t+1}}\left[-2\gamma M_\times^{t+1} H^{t+1} + 2\eta_H \delta P_H^t H^{t+1}\right] = 0.$$

The subsequent weight and tensor updates do not change the derivative with respect to H, so

$$D_H L_{\rho_s}(\Theta^{t+1}) = -2\gamma M_\times^{t+1} H^{t+1}.$$

Hence

$$\begin{aligned}\Pi_{H^{t+1}}\left(D_H L_{\rho_s}(\Theta^{t+1})\right) &= -2\eta_H \Pi_{H^{t+1}}(\delta P_H^t H^{t+1}), \\ \left\|\Pi_{H^{t+1}}\left(D_H L_{\rho_s}(\Theta^{t+1})\right)\right\|_F &\le 2\eta_H \|\delta P_H^t\|_F \longrightarrow 0.\end{aligned} \tag{L1.6}$$

Here we used nonexpansiveness of $\Pi_{H^{t+1}}$, $\|H^{t+1}\|_2 = 1$, and (L1.2).

The ordinary Lagrangian. Set

$$\widehat{\mathcal{Y}}^{s,t+1} = \mathcal{Y}^s + \rho_s \mathcal{R}(\Theta^{t+1}), \qquad L(\Theta, \mathcal{Y}) = \mathcal{Q}(\Theta) + \langle \mathcal{Y}, \mathcal{R}(\Theta)\rangle.$$

For every smooth block X, the chain rule gives

$$\begin{aligned}D_X L_{\rho_s}(\Theta^{t+1}) &= D_X \mathcal{Q}(\Theta^{t+1}) \\ &\quad + [D_X \mathcal{R}(\Theta^{t+1})]^*\left[\mathcal{Y}^s + \rho_s \mathcal{R}(\Theta^{t+1})\right] \\ &= D_X L(\Theta^{t+1}, \widehat{\mathcal{Y}}^{s,t+1}).\end{aligned}$$

Here * denotes the adjoint of the derivative of $\mathcal{R}$, and the candidate multiplier is held fixed when differentiating the ordinary Lagrangian.

The stopping condition. With all ordinary-Lagrangian derivatives evaluated at $(\Theta^{t+1}, \widehat{\mathcal{Y}}^{s,t+1})$, the prescribed stopping residual is

$$\mathcal{S}_s^{t+1} = \max\Big\{ \max_{1\le v\le V} \|\Pi_{F_v^{t+1}}(D_{F_v}L)\|_F, \max_{1\le v\le V} \|\Pi_{R_v^{t+1}}(D_{R_v}L)\|_F, \quad \|\Pi_{H^{t+1}}(D_H L)\|_F \Big\}.$$

The gradient identity above, (L1.4), (L1.5), and (L1.6) imply $\mathcal{S}_s^{t+1} \to 0$, since V is finite. For each fixed $\varepsilon_s > 0$, there is therefore a finite t_0 such that $\mathcal{S}_s^{t+1} \le \varepsilon_s$ for all $t \ge t_0$. Hence the inner loop terminates after finitely many complete sweeps. □

Lemma 2. *Let $r > 1$, $\lambda \ge 0$, and $\rho_s > 0$. At outer iteration s, let $\Theta^{s+1} = \Theta^{s,t_s}$, with $t_s \ge 1$, be the accepted point after a complete sweep. Suppose that its blocks satisfy the Stiefel, orthogonal, and simplex constraints and that $\mathcal{S}_s^{t_s} \le \varepsilon_s$, with all residuals evaluated after the final $\mathcal{J}$-update. Assume that the final $\boldsymbol{\alpha}$- and $\mathcal{J}$-updates are computed exactly in the prescribed sweep order, including when view energies vanish. Set*

$$\mathcal{Y}^{s+1} = \mathcal{Y}^s + \rho_s\big(\widetilde{\mathcal{F}}^{s+1} - \mathcal{J}^{s+1}\big).$$

At $(\Theta^{s+1}, \mathcal{Y}^{s+1})$, each F_v-, R_v-, and H-block stationarity residual of the split problem is at most ε_s. The $\boldsymbol{\alpha}$- and $\mathcal{J}$-block stationarity conditions hold exactly; in particular,

$$\mathcal{Y}^{s+1} \in \lambda \partial \|\mathcal{J}^{s+1}\|_{\mathrm{TNN}}.$$

Proof. Write $\Theta = \Theta^{s+1}$ and suppress the superscripts of its components. With $\mathcal{R} = \widetilde{\mathcal{F}} - \mathcal{J}$ and $\mathcal{R}_v = F_v R_v - \mathcal{J}_v$, the Lagrangians are

$$\begin{aligned} L(\Theta, \mathcal{Y}) &= \mathcal{Q}(\Theta) + \langle \mathcal{Y}, \mathcal{R}(\Theta)\rangle, \\ L_{\rho_s}(\Theta; \mathcal{Y}^s) &= L(\Theta, \mathcal{Y}^s) + \frac{\rho_s}{2}\|\mathcal{R}(\Theta)\|_F^2. \end{aligned}$$

Here $\mathcal{Q}$ is the original objective, without the block proximal penalties. Write $L_{\rho_s}(\Theta) = L_{\rho_s}(\Theta; \mathcal{Y}^s)$. Multipliers are held fixed when differentiating either Lagrangian.

Let $\mathcal{X} = (F_1, \dots, F_V, R_1, \dots, R_V, H)$ denote the manifold blocks, with $\mathcal{C}_{F_v} = \mathcal{C}_H = \mathrm{St}(n, c)$ and $\mathcal{C}_{R_v} = \mathrm{O}(c)$. Let $N_{\mathcal{C}_X}(X)$ be the Euclidean normal space and $N_{\Delta_V}(\boldsymbol{\alpha})$ the convex normal cone to the simplex. Define

$$\begin{aligned} \mathcal{K}_s = \max_{X\in\mathcal{X}} & \operatorname{dist}\big(0, D_X L(\Theta, \mathcal{Y}^{s+1}) + N_{\mathcal{C}_X}(X)\big) \\ & + \operatorname{dist}(0, D_{\boldsymbol{\alpha}}\mathcal{Q} + N_{\Delta_V}(\boldsymbol{\alpha})) \\ & + \operatorname{dist}\big(0, \lambda\partial\|\mathcal{J}\|_{\mathrm{TNN}} - \mathcal{Y}^{s+1}\big) \\ =: \mathrm{I} + \mathrm{II} + \mathrm{III}. \end{aligned} \tag{L2.1}$$

Distances use the Frobenius norm for matrices and tensors and the Euclidean norm for weights.

Bound for I. The residual differential is $\mathrm{d}\mathcal{R}_v[A, B, C] = AR_v + F_v B - C$, where A, B, C perturb $F_v, R_v, \mathcal{J}_v$, respectively. Thus the chain rule and $\mathcal{Y}_v^s + \rho_s\mathcal{R}_v = \mathcal{Y}_v^{s+1}$ give

$$\begin{aligned} D_{F_v}L_{\rho_s} &= D_{F_v}\mathcal{Q} + \mathcal{Y}_v^{s+1}R_v^\top = D_{F_v}L(\Theta, \mathcal{Y}^{s+1}), \\ D_{R_v}L_{\rho_s} &= F_v^\top\mathcal{Y}_v^{s+1} = D_{R_v}L(\Theta, \mathcal{Y}^{s+1}), \\ D_H L_{\rho_s} &= -2\gamma M_\times H = D_H L(\Theta, \mathcal{Y}^{s+1}). \end{aligned} \tag{L2.2}$$

Here we used the independence of $\mathcal{Q}$ from R_v and the symmetry of $M_\times$.

For $X \in \mathrm{St}(p, c)$, the tangent projection under the embedded Euclidean metric is $\Pi_X(Z) = Z - X\,\mathrm{sym}(X^\top Z)$, and the normal space is $\{X\Lambda : \Lambda^\top = \Lambda\}$ (Boumal, 2023, Sec. 7.3). Here $\mathrm{sym}(A) = (A + A^\top)/2$. The orthogonal decomposition $Z = \Pi_X(Z) + X\,\mathrm{sym}(X^\top Z)$ therefore gives

$$\operatorname{dist}\big(0, Z + N_{\mathrm{St}(p,c)}(X)\big) = \|\Pi_X(Z)\|_F. \tag{L2.3}$$

Taking $p = n$ for F_v, H and $p = c$ for $R_v \in \mathrm{O}(c) = \mathrm{St}(c, c)$, (L2.2) and (L2.3) give

$$\mathrm{I} = \max_{X\in\mathcal{X}} \|\Pi_X(D_X L_{\rho_s})\|_F = \mathcal{S}_s^{t_s} \le \varepsilon_s. \tag{L2.4}$$

The stopping residual is evaluated at this same primal point with candidate multiplier $\widehat{\mathcal{Y}}^{s,t_s} = \mathcal{Y}^{s+1}$.

Bound for II. Since $r > 1$, the weighted energy admits the continuously differentiable extension $\sum_v |\alpha_v|^r E_v(F_v)$. Exact minimization over Δ_V gives the first-order condition

$$0 \in D_{\boldsymbol{\alpha}}\mathcal{Q} + N_{\Delta_V}(\boldsymbol{\alpha}) = \big(r\alpha_v^{r-1}E_v(F_v)\big)_{v=1}^{V} + N_{\Delta_V}(\boldsymbol{\alpha}).$$

This condition also holds when weights or energies vanish, since $|a|^r$ has derivative zero at $a = 0$. The subsequent $\mathcal{J}$-update changes neither $\boldsymbol{\alpha}$ nor $E_v(F_v)$, so $\mathrm{II} = 0$ at the accepted point.

Bound for III. The tensor update is the proximal minimization

$$\mathcal{J} \in \arg\min_{\mathcal{Z}} \left\{ \lambda\|\mathcal{Z}\|_{\mathrm{TNN}} + \frac{\rho_s}{2}\left\|\mathcal{Z} - \left(\widetilde{\mathcal{F}} + \frac{\mathcal{Y}^s}{\rho_s}\right)\right\|_F^2 \right\}.$$

Its optimality condition (Parikh & Boyd, 2014, Sec. 2.3) is

$$\begin{aligned} 0 &\in \lambda\partial\|\mathcal{J}\|_{\mathrm{TNN}} - \mathcal{Y}^s - \rho_s\big(\widetilde{\mathcal{F}} - \mathcal{J}\big) \\ &= \lambda\partial\|\mathcal{J}\|_{\mathrm{TNN}} - \mathcal{Y}^{s+1}. \end{aligned}$$

Thus $\mathrm{III} = 0$. Substitution into (L2.1), together with (L2.4) and $\mathrm{II} = 0$, gives $\mathcal{K}_s \le \varepsilon_s$ and the stated exact weight and tensor stationarity conditions. □

Lemma 3. *Under the setting of Lemma 2, suppose that $\mathcal{Y}^0 = 0$ and that $\rho_s \ge \rho_0 > 0$ for all $s \ge 0$. Then, for every $s \ge 0$,*

$$\begin{aligned} &\|\mathcal{Y}^{s+1}\|_F \le \lambda\sqrt{c}, \\ &\|\widetilde{\mathcal{F}}^{s+1}\|_F = \sqrt{Vc}, \\ &\|\mathcal{J}^{s+1}\|_F \le \sqrt{Vc} + \frac{2\lambda\sqrt{c}}{\rho_0}. \end{aligned} \tag{L3.1}$$

Consequently, every infinite sequence of accepted primal–dual iterates $\{(\Theta^{s+1}, \mathcal{Y}^{s+1})\}_{s\ge 0}$ is bounded and has an accumulation point.

Proof. Rearranging the multiplier update gives

$$\mathcal{J}^{s+1} = \widetilde{\mathcal{F}}^{s+1} - \frac{\mathcal{Y}^{s+1} - \mathcal{Y}^s}{\rho_s}.$$

Hence

$$\begin{aligned} \|\mathcal{J}^{s+1}\|_F &\le \|\widetilde{\mathcal{F}}^{s+1}\|_F + \frac{\|\mathcal{Y}^{s+1}\|_F}{\rho_s} + \frac{\|\mathcal{Y}^s\|_F}{\rho_s} \\ &=: \mathrm{I} + \mathrm{II} + \mathrm{III}. \end{aligned} \tag{L3.2}$$

Bound for I. Since $(F_v^{s+1})^\top F_v^{s+1} = I_c$ and $(R_v^{s+1})^\top R_v^{s+1} = I_c$,

$$\begin{aligned} \mathrm{I}^2 &= \sum_{v=1}^{V} \|F_v^{s+1}R_v^{s+1}\|_F^2 \\ &= \sum_{v=1}^{V} \mathrm{Tr}\big[(R_v^{s+1})^\top (F_v^{s+1})^\top F_v^{s+1} R_v^{s+1}\big] = Vc. \end{aligned}$$

Thus $\mathrm{I} = \sqrt{Vc}$.

Bound for II. For $\mathcal{B} \in \mathbb{R}^{n\times c\times V}$, let $\widehat{\mathcal{B}}$ denote its unnormalized discrete Fourier transform along the third mode. The TNN normalization, Cauchy–Schwarz inequality, and Parseval identity give

$$\begin{aligned} \|\mathcal{B}\|_{\mathrm{TNN}} &= \frac{1}{V}\sum_{\ell=1}^{V} \|\widehat{\mathcal{B}}^{(\ell)}\|_* \\ &\le \frac{\sqrt{c}}{V}\sum_{\ell=1}^{V} \|\widehat{\mathcal{B}}^{(\ell)}\|_F \le \sqrt{\frac{c}{V}\sum_{\ell=1}^{V} \|\widehat{\mathcal{B}}^{(\ell)}\|_F^2} = \sqrt{c}\,\|\mathcal{B}\|_F. \end{aligned} \tag{L3.3}$$

If $\mathcal{G} \in \partial\|\mathcal{A}\|_{\mathrm{TNN}}$, the subgradient and triangle inequalities, together with (L3.3), give, for every real tensor $\mathcal{B}$ of the same size,

$$\begin{aligned}\langle \mathcal{G}, \mathcal{B}\rangle &\le \|\mathcal{A}+\mathcal{B}\|_{\mathrm{TNN}} - \|\mathcal{A}\|_{\mathrm{TNN}} \\ &\le \|\mathcal{B}\|_{\mathrm{TNN}} \le \sqrt{c}\,\|\mathcal{B}\|_F.\end{aligned}$$

Taking $\mathcal{B} = \mathcal{G}$ yields $\|\mathcal{G}\|_F^2 \le \sqrt{c}\,\|\mathcal{G}\|_F$, and hence $\|\mathcal{G}\|_F \le \sqrt{c}$.

By Lemma 2,

$$\mathcal{Y}^{s+1} \in \lambda\partial\|\mathcal{J}^{s+1}\|_{\mathrm{TNN}}.$$

For $\lambda > 0$, the preceding bound applies to $\mathcal{Y}^{s+1}/\lambda$; for $\lambda = 0$, the inclusion gives $\mathcal{Y}^{s+1} = 0$. Therefore,

$$\|\mathcal{Y}^{s+1}\|_F \le \lambda\sqrt{c}, \qquad \mathrm{II} \le \frac{\lambda\sqrt{c}}{\rho_0}, \qquad s \ge 0. \tag{L3.4}$$

Bound for III. For $s = 0$, $\mathcal{Y}^0 = 0$ gives III $= 0$. For $s \ge 1$, applying (L3.4) at the preceding outer iteration gives

$$\mathrm{III} = \frac{\|\mathcal{Y}^s\|_F}{\rho_s} \le \frac{\lambda\sqrt{c}}{\rho_0}.$$

Substituting these three bounds into (L3.2) proves (L3.1).

The remaining blocks satisfy

$$\sum_{v=1}^{V}\|F_v^{s+1}\|_F^2 = \sum_{v=1}^{V}\|R_v^{s+1}\|_F^2 = Vc, \qquad \|H^{s+1}\|_F^2 = c, \qquad \|\boldsymbol{\alpha}^{s+1}\|_2 \le 1.$$

Together with (L3.1), these bounds show that the accepted primal–dual sequence is bounded in a finite-dimensional Euclidean space. It therefore has a convergent subsequence. □

Lemma 4. *Under the setting of Lemma 3, suppose that the prescribed adaptive penalty rule uses fixed parameters $0 < \theta < 1$ and $\kappa > 1$. Then every infinite sequence of accepted outer iterates satisfies*

$$\left\|\widetilde{\mathcal{F}}^{s+1} - \mathcal{J}^{s+1}\right\|_F \longrightarrow 0 \qquad \text{as } s \to \infty.$$

Proof. Write $\mathfrak{r}_{s+1} = \|\widetilde{\mathcal{F}}^{s+1} - \mathcal{J}^{s+1}\|_F$. The multiplier update and Lemma 3, together with $\mathcal{Y}^0 = 0$, give, for every $s \ge 0$,

$$\mathfrak{r}_{s+1} = \frac{\|\mathcal{Y}^{s+1} - \mathcal{Y}^s\|_F}{\rho_s} \le \frac{\|\mathcal{Y}^{s+1}\|_F + \|\mathcal{Y}^s\|_F}{\rho_s} \le \frac{2\lambda\sqrt{c}}{\rho_s}. \tag{L4.1}$$

The first penalty update sets $\rho_1 = \rho_0$. Along an infinite accepted run, subsequent updates satisfy

$$\rho_{s+1} = \begin{cases}\rho_s, & \mathfrak{r}_{s+1} \le \theta\mathfrak{r}_s,\\ \kappa\rho_s, & \mathfrak{r}_{s+1} > \theta\mathfrak{r}_s,\end{cases} \qquad s \ge 1. \tag{L4.2}$$

If the penalty parameter increases infinitely often, each increase multiplies ρ_s by the fixed factor $\kappa > 1$, while all other updates leave it unchanged. Thus $\rho_s \to \infty$, and (L4.1) gives

$$0 \le \mathfrak{r}_{s+1} \le \frac{2\lambda\sqrt{c}}{\rho_s} \longrightarrow 0.$$

Otherwise, the penalty parameter increases only finitely many times. Choose $s_0 \ge 1$ such that $\rho_{s+1} = \rho_s$ for every $s \ge s_0$. Since $\rho_s > 0$ and $\kappa > 1$, the update rule (L4.2) then implies $\mathfrak{r}_{s+1} \le \theta\mathfrak{r}_s$ for every $s \ge s_0$. Repeated application yields

$$0 \le \mathfrak{r}_{s+1} \le \theta^{\,s-s_0+1}\mathfrak{r}_{s_0} \longrightarrow 0, \qquad s \to \infty,$$

because $0 < \theta < 1$. □

C Dataset Details

- **MHEALTH.** Human activity data with chest, arm, and ankle motion views. Source: `https://archive.ics.uci.edu/dataset/319/mhealth+dataset`.
- **DSADS.** Human activity data with five body-worn IMU views. Source: `https://archive.ics.uci.edu/dataset/256/daily+and+sports+activities`.
- **Tactile12.** Tactile texture data with acceleration, angular-rate, and magnetic-field views. Source: `https://data.mendeley.com/datasets/n666tk4mw9/1`.
- **PASTIS-R.** Crop-parcel data with Sentinel-1 ascending, Sentinel-1 descending, and Sentinel-2 optical views. Source: `https://github.com/VSainteuf/pastis-benchmark`.
- **HPC-ODA.** HPC application data with CPU performance, memory, network, CPU/OS, and power/temperature views. Source: `https://zenodo.org/records/4671477`.
- **NinaPro DB5-S1.** Hand-gesture data with two Myo EMG views and one glove view. Source: `https://ninapro.hevs.ch/instructions/DB5.html`.